REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
	Final Technical Report	01/June/01 - 31/Dec/06

4. TITLE AND SUBTITLE

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield

5a. CONTRACT NUMBER

5b. GRANT NUMBER

DAAD19-01-1-0621

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Peter Hancock

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Central Florida
4000 Central Florida Blvd.
Orlando, FL 32816

8. PERFORMING ORGANIZATION REPORT NUMBER

42244.44-LS-MUR

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
PO BOX 12211
Durham, NC 27709

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

N/A

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	2950	**19b. TELEPHONE NUMBER** *(include area code)*
UU	UU	UU			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

University of Central Florida

Department of Psychology

To: Dr. Elmar T. Schmeisser
From: Peter A. Hancock
Re: MURI Final Report DVD
Date: 28 March 2007

Dear Dr. Schmeisser,

I am very glad to present to you the MURI Final Report, which is here provided on a single DVD. We are, of course, happy to send copies to whomever you might require and have already sent one to Dr. Allender as per your e-mail. We are very proud of the effort and would like to give it the widest possible dissemination. As always, your guidance and direction is very much appreciated.

Regards,

Peter Hancock

42244

Partnership II • 3100 Technology Pkwy, Ste 337 • Orlando, FL 32826-0544
Telephone 407-823-1492 • email hfesras2@mail.ucf.edu

An Equal Opportunity and Affirmative Action Institution

University of Central Florida

Multi-Disciplinary University Research Initiative

MURI-OPUS

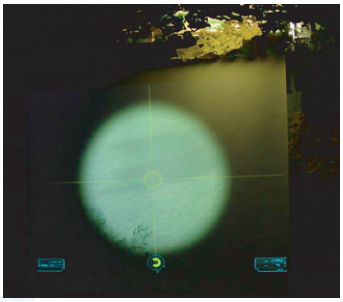
Operator Performance Under Stress

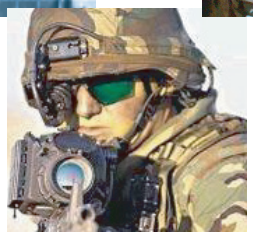
www.mit.ucf.edu

VOLUME FIVE











2005 Executive Summary

Table of Contents

NCRADA

A Collaborative Partnership between:

University of Central Florida

Contact Information:

University of Central Florida • MIT² Laboratory

3100 Technology Pkwy, Suite 337 • Partnership II • Orlando, Florida 32826-0544

http://www.mit.ucf.edu

Table of Contents

MURI Divisions

University of Central Florida (Peter Hancock)

Supported Work

Information Processing Changes Following Extended Stress *Refereed Publication* Harris, W.C., Hancock, P.A., & Harris, S.C. (2005) *Military Psychology*, *17* (2), 115-128.

A Meta-Analytic Review of the Effects of Temperature on Performance *Abstract* American Psychological Association Division 21/Division 19 and Human Factors and Ergonomics Society Potomac Chapter Annual Symposium on Applied Experimental Research. *Ross, J.M., Szalma, J.L., Oron-Gilad, T., Hancock, P.A. (2005).*

Perception of Morphed Objects: Generating Psychophysical Functions for the Evolution of a Shape *Presentation* American Psychological Association Division 21/Division 19 and Human Factors and Ergonomics Society Potomac Chapter Annual Symposium on Applied Experimental Research. *Morgan, J.F., Oron-Gilad, T., Hancock, P.A., & Szalma, J.L. (2005).*

Stress, Workload, and the Detection of Threat *Abstract* Bugging Bugs and Brains: Communication and Applications from Cellular internals to Mental Contents, A Workshop/Symposium sponsored by the U S Army Research Office *Szalma, J.L. & Hancock, P.A. (2005)*

Individual Differences in Workload, Stress, and Coping in Police Officers Engaged in Shooting Tasks *Poster* Eighth International Symposium on Human Factors in Organizational Design and Management *Szalma, J.L., Oron-Gilad, T., & Hancock, P.A. (2005).*

Incorporating Individual Differences into the Adaptive Automation Paradigm *Poster* Eighth International Symposium on Human Factors in Organizational Design and Management *Oron-Gilad, T., Szalma, J.L., Thropp, J.E., & Hancock, P.A. (2005).*

Time, Emotion, and the Limits to Human Information Processing and the Organizations they Form *Poster* Eighth International Symposium on Human Factors in Organizational Design and Management *Hancock, P.A., Szalma, J.L. & Oron-Gilad, T. (2005).*

Operators' Time Perception under Stress *Paper* Proceedings of the International Symposium on Aviation Psychology *Nayeem, R.V., Oron-Gilad, T., & Hancock, P.A. (2005)*

Pilots' Performance under Stress *Presentation* International Symposium on Aviation Psychology *Nayeem, R.V., Oron-Gilad, T., & Hancock, P.A. (2005)*

The Driving Simulator as a Workload Research Tool *Abstract* Driving Simulator Conference North America 2005 *Morgan, J.F., Oron-Gilad, T., & Hancock, P.A. (2005)* .

Road Environment and Passive Fatigue *Poster* 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. *Oron-Gilad, T. & Hancock, P.A. (2005).*

Road Environment and Passive Fatigue *Abstract* 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. *Oron-Gilad, T. & Hancock, P.A. (2005).*

Comparison of Spatial and Temporal Discrimination Performance across Various Difficulty Levels *Proceedings* Human Factors and Ergonomics Society's 49th Annual Meeting. *Thropp, J.E., Szalma, J.L., & Hancock, P.A. (2005)* in press.

Why and How HFE Professionals Can Better Use Theory (Metatheory Included: Some Assembly Required) *Proceedings* Human Factors and Ergonomics Society 49th Annual Meeting. *Koltko-Rivera, M. & Hancock, P.A. (2005)* in press.

On the Relationship between Workload and Performance *Submitted Manuscript* Oron-Gilad, T., Szalma, J.L., Stafford, S.C., & Hancock, P.A. (2005).

Training for Vigilance: Using Predictive Power to Evaluate Feedback Effectiveness. *Refereed Publication* Szalma, J.L., Hancock, P.A., Warm, J.S., Dember, W.N., & Parsons, K.S. (2005) Training for Vigilance: Using Predictive Power to Evaluate Feedback Effectiveness. *Human Factors*, in press.

Are All Teams Created Equal? An examination of teamwork communication across teams. *Submitted Manuscript* Priest, H. A., Burke, C. S., Guthrie, J., Salas, E. & Bowers, C. A. in review, *Human Factors (2005).*

Neuropsychological Assessments in Extreme Environments *Submitted Manuscript* Lowe, M., Harris, W., Kane, R. Banderet, L.E., Levinson, D., & Reeves, D.L. (2005).

Cognitive Change in Special Forces Personnel Following Stressful Survival Training *Proceedings* Human Factors and Ergonomics Society 49th Annual Meeting *Harris, W.C., Hancock, P.A. & Morgan III, C.A. (2005)* in press.

Testing the Efficiency and Effectiveness of Soldier Performance Under Increasing Task Demand *Proceedings* Human Factors and Ergonomics Society 49th Annual

Meeting *Burke, K.A., Szalma, J.L., Oron-Gilad, T., Duley, A., & Hancock, P.A. (2005)* in press.

Evaluation of Threat by Police Officers: Initial Findings *Proceedings* Human Factors and Ergonomics Society 49th Annual Meeting *Delos Santos, K.A., Stafford, S.C., Oron-Gilad, T., & Hancock, P.A. (2005)* in press.

Training for Vigilance: The Effect of KR Format and Dispositional Optimism and Pessimism on Performance and Stress. *Refereed Publication* Szalma, J.L., Hancock, P.A., Dember, W.N., & Warm, J.S. (2005) Training for Vigilance: The Effect of KR Format and Dispositional Optimism and Pessimism on Performance and Stress. *British Journal of Psychology*, in press.

Facilitated Work

Remotely Operated Vehicles from the Top Down and the Bottom Up *Book Chapter Human Factors in Human Performance and Cognitive Engineering Research Series*. Oron-Gilad, T., Chen, J.Y.C., & Hancock, P.A. *(2005)* in press.

Hedonomics: The Power of Positive and Pleasurable Ergonomics *Refereed Publication* Hancock, P.A., Pepe, A.A., & Murphy, L.L. (2005) Hedonomics: The Power of Positive and Pleasurable Ergonomics *Ergonomics in Design* Winter pp. 8-14.

Police Officers Seat Belt Use While on Duty *Refereed Publication* Oron-Gilad, T., Szalma, J.L., Stafford, S.C., & Hancock, P.A. (2005) *Transportation Research Part F, 8*, pp. 1-18.

To Honor Wickens and Re-Stack His Boxes *Paper* Applied Attention: From Theory to Practice Conference *Hancock, P.A., Oron-Gilad, T., & Szalma, J.L. (2005).*

To Honor Wickens and Re-Stack His Boxes *Poster* Applied Attention: From Theory to Practice Conference *Hancock, P.A., Oron-Gilad, T., & Szalma, J.L. (2005).*

The Role of Hedonomics in the Future of Industry, Service, and Product Design: Panel Overview *Panel Session* Human Factors and Ergonomics Society 49th Annual Meeting *Chair- Hancock, P.A., Co-Chair- Oron-Gilad, T. Panelists: Oron-Gilad, T., Szalma, J.L., Khalid, H.M., Meyer, J., & Hancock, P.A. (2005)* in press.

Exploiting the Human in Human Machine Systems *Presentation* 11th International Conference on Human-Computer Interaction *Oron-Gilad, T., Hancock, P.A., & Szalma, J.L. (2005).*

Are We Ready to Consider Individual Difference in Human Capabilities in Our Workplace Designs? *Panel Session* Human Factors and Ergonomics Society 49th Annual Meeting *Chair- Gielo-Perczak, K., Co-chair-Karwowski, W.. Panelists- Hancock, P.a., Gielo-Perczak, K., Karwowski, W., Marras, W.S. Parasuraman, R.* (2005) in press.

Linköping Institute of Technology (Kip Smith)

Supported Work

Paintball 2: A Study of Remote-Controlled Tactile Stimulation for the Simulation of Leader Presence. *Proceedings* 49th Annual Meeting of the Human Factors and Ergonomics Society. *Smith, K. & Liberg, M. (2005).* In press.

HUD Use Improves Landing Performance in Restricted Visibility. *Submitted Manuscript Goteman, Ö, Smith, K., & Dekker, S. (2005)* in revision.

Progress Report – 01-2005

Progress Report – 04-2005

George Mason University (Raja Parasuraman)

Supported Work

Effects of imperfect automation on decision making in a simulated command and control task. *Refereed Publication – in revision* Human Factors, revision under review. *Rovira, E., McGarry, K., & Parasuraman, R. (2005).*

Adaptive change in the type of automation support reduces the cost of imperfect decision aids in a simulated battlefield engagement task. *Proceedings* 49th Annual Human Factors & Ergonomics Society *McGarry, K., & Parasuraman, R. (2005)* in press.

Progress Report – 04/2005

MILITARY PSYCHOLOGY, 2005, *17*(2), 115–128
Copyright © 2005, Lawrence Erlbaum Associates, Inc.

Information Processing Changes Following Extended Stress

Wayne C. Harris, P. A. Hancock, and Scot C. Harris
Department of Psychology and Institute for Simulation and Training
University of Central Florida

The psychological states and cognitive performance capacities of military participants were assessed prior to and following 1 week of field training at a Navy Survival, Evasion, Resistance, and Escape (SERE) School. The effect of this exposure resulted in significant cognitive performance deterioration and an increase in reported levels of subjective discomfort. Simple Reaction Time increased, and the capacity to perform information manipulation tasks such as Spatial Processing and Code Substitution deteriorated. Logical Reasoning proved the most susceptible of all tasks to the effects of the weeklong stressful exposure. These various effects were evident only in interaction with time on task. In the testing session following the stressful field exposure, participants' immediate response was at a level similar to their pretraining baseline. However, unlike their preexposure session, performance in the later condition declined as the session progressed. This pattern suggests that immediately following the stress exposure, participants are able to sustain their normal performance but only for a limited period of time. Because the duration of any one single task, each being less than 1 min, was very brief, the results indicated that this time-based degradation was fairly rapid. Overall results indicate, therefore, that for stress assessment, Simple Reaction Time task may be usefully and pragmatically diagnostic of stress-induced deterioration. Our results further indicate that cognitive performance decrements are associated with subjective report of decreased vigor and increased difficulty in concentrating. In addressing the issue of sustained performance capacity and resilience following extended stress exposure, we seek to facilitate one goal—a modern objective force that is responsive to rapid changes in mission profile and operations. These findings indicate that crucial information-processing response measures help to achieve this aim.

Do not print

Requests for reprints should be sent to Wayne C. Harris, Department of Psychology and Institute for Simulation and Training, University of Central Florida, Orlando, FL 32816. E-mail: wcharris@qwest.net

In combat, soldiers experience sleep loss, high levels of physical and psychological discomfort, prolonged periods of heightened vigilance, and extreme danger (Hancock & Hoffman, 1997). Such battlefield conditions can completely incapacitate some soldiers (Mareth & Brooker, 1982), but the response variation of soldiers who continue to fight is more difficult to estimate. Performance disruption produced by stress can be inferred from reports that a large proportion of soldiers fail to fire their weapons in combat (Marshall, 1947) and the demonstration of impaired performance in simulated battlefield conditions (Villoldo & Tarno, 1984). Recognition of that and of allied findings means that the concern for the effect of stress on military personnel is not a new one. However, its importance is increasing in proportion to the complexity of equipment being developed to enhance war-fighter capability. Preparing for the high-tech battlefield therefore requires knowledge of information-processing tasks or task components that promise to be most vulnerable to stress (Orasanu & Backer, 1996). Our purpose here is to evaluate the change in such capacities under the effects of a realistic, operational stress.

Military personnel clearly exhibit performance variation during times of real stress (Belland & Bissell, 1994). Berkum (1964) showed that there are disruptive effects in response to just the perception of danger. Performance decrements in combat would be predicted because military decisions involve such a wide spectrum of information-processing components (Wickens & Flach, 1988), and stress impairs each of these different information-processing stages (see Hancock, 1986; Wickens, 1996). However, such stress-induced cognitive changes that have been proposed as the basis of reduced operational performance have been difficult to detect in general (Callister, Percival, & Retzlaff, 1999; Elsmore, Naitoh, & Linnville, 1992; Slaven & Windle, 1999).

Estimating the cognitive changes that occur in operational settings will always be difficult. Combat involves highly stressful incidents, embedded in extended periods of quiescence that still represent a chronic stress. The majority of soldier decisions are not made during actual combat but in comparatively less stressful, normal operating conditions. However, the quality of even these decisions is crucial to operational effectiveness. It is anathema to administer assessment instruments during an infantry engagement, but it is possible to collect data that can be used to model and predict cognitive performance change in the operational environment. In our study we sought to examine this issue by administering a battery of performance tasks immediately following stressful conditions specifically constructed to represent as closely as possible those present in actual operational settings. Thus, the specific goals of our work were (a) to assess the type and amount of cognitive change that follow a period of sustained physical and psychological stress, (b) to measure the subjective change that occurred during this period, and (c) to determine the relationship between subjective and cognitive performance change.

Do Not Copy

EXPERIMENTAL METHOD

Experimental Participants

Thirty-five active duty Navy and Marine Corps personnel enrolled in the Navy Survival, Evasion, Resistance, and Escape (SERE) School in Brunswick, Maine, acted as participants in the study. They were not recompensed in any way for their voluntary participation.

Assessment Instruments

The performance of complex tasks is often conceived as dependent on the person's ability to complete a sequence of simpler operations (Wickens & Flach, 1988). In consequence, a cognitive battery was selected that included tasks that represent critical components of complex real-world performance. The abilities to attend, perceive, abstract, and process information and to access information from memory were the prime components of information processing chosen. The effects of stress on information processing should be reflected in changes in one or more of these respective components. The Automated Neurological Assessment Metric (ANAM) battery was developed to assess such components of cognitive performance (Reeves et al., 1991). The ANAM is a computerized battery of cognitive functioning tests adapted from the Walter Reed Performance Assessment Battery (Englund et al., 1987), and respective subtests are available that assess the cited information-processing stages (Englund et al., 1987). In our study, subtests from the ANAM were used; the subtests assess cognitive abilities that have been proposed to comprise the information-processing components of complex tasks, including reaction time, stimulus encoding, memory, and the manipulation of spatial and symbolic information. The subtests were also selected based on their appropriateness for testing in sustained operations settings (Reeves et al., 1991). Selected tests included Simple Reaction Time (SRT), Code Substitution (CDS), Spatial Processing (SPD), Logical Reasoning (LRS), two memory set tasks (i.e., a Sternberg memory task [Sternberg, 1969] with a two-item memory set [ST2] and a six-item memory set [ST6]), and the Continuous Performance Task (CPT).

The experimental design entailed administration of the ANAM three times before field training and a comparison of the third and last prestress administration with immediate postfield training scores. Benedetto, Harris, and Goernert (1995) found that for repeated ANAM administration, scores improved rapidly during the first three administrations, after which time there were only small improvements. A classroom training session was included to maximize the effect of the present training regime. A particular concern was the effect of the weeklong exposure between the third and fourth ANAM administrations. Benedetto et al. only examined performance during repeated testing when assessments occurred during a single

session. To estimate the expected ANAM score change between Sessions 3 and 4 when 1 week separated the sessions, 15 student members of the Army Reserve Officer Training Corps completed the same testing regime as the study participants. Accuracy and response time scores across Sessions 3 and 4 did not differ significantly, multivariate analysis of variance (MANOVA), $F(7, 8) = 0.47$, $p < .84$, and $F(7, 8) = 2.18$, $p < .15$, respectively. There was a trend toward faster responding in the more complex tasks (LRS, CDS, and SPD) during the fourth session. These results are consistent with all previous findings that the majority of improvement during repeated administration of the ANAM occurs during the second and third administrations but that small changes may be anticipated in the more complex tasks. It is evident from the test performed that this pattern persists when the third and fourth sessions are separated by 1 week. As a result of this companion evaluation, we ensured that changes in performance observed in the survival school participants were due to the stress exposure, not simply the weeklong hiatus in the training schedule.

Subjective state was assessed using the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1992), and the Sustained Operations Assessment Profile (SOAP; Retzlaff, King, Marsh, & French, 1997). The POMS comprises 65 adjectives, and participants were directed to indicate on a 5-point scale the degree to which each adjective was consistent with their current state. The POMS measures six specific mood states: Tension–Anxiety, Depression–Dejection, Anger–Hostility, Vigor–Activity, Fatigue–Inertia, and Confusion–Bewilderment. Approximately 5 min are required to complete the POMS. The SOAP is a list of 90 short phrases. Participants indicate the degree to which each item matches their current feelings on a 5-point scale. The SOAP scales are Poor Concentration, Boredom, Slowed Reactions, Anxiety, Depression, Irritability, Fatigue/Low Energy, Poor Sleep, Work Frustration, and Physical Discomfort.

Do not copy

Experimental Procedure

After signing the informed consent document, participants completed an ANAM training session, which included classroom instruction that summarized the assessment materials and two practice sessions. The training battery included on-screen instructions and comprised 10 trials per task. After training, participants completed the ANAM, the POMS, and the SOAP. Two SERE classes were tested. Prefield assessment of the first SERE class of 19 occurred at 1600 hr the day before they began field training. The second class of 16 completed the battery in the morning immediately before being transported to the training site. SERE students then completed the 7-day field-training exercise, which included survival training and an exercise that simulated being isolated in hostile territory. The training presented many of the elements of military operations, including physical discomfort, sleep loss, and high levels of perceived threat. Training encompasses those basic

skills necessary for worldwide survival: expedite search and rescue efforts; evade capture by hostile forces; resist interrogation, exploitation, and indoctrination; and escape from detention by enemy forces in accordance with Department of Defense Directive 1300.7. Participants completed the ANAM battery, the POMS, and the SOAP immediately after they disembarked from the transport, which had brought them from the training site. The postfield training assessments were obtained at 1600 hr.

EXPERIMENTAL RESULTS

Cognitive performance and subjective state were measured before and after 1 week of SERE training. Assessment batteries contained multiple subscales; therefore, pre- and postfield training scores were initially compared using a MANOVA employing the Wilks's lambda criterion to determine whether pre- and posttraining assessments differed. Contingent on the outcome of the overall MANOVA significance level, individual scale changes were examined to determine their contribution to any observed differences. An analysis was then performed to examine the relationship between subjective state changes and cognitive performance changes.

Do not Copy

Cognitive Performance

Analysis showed that accuracy changed significantly during the week of field training, MANOVA, $F(6, 24) = 5.56$, $p < .001$. Univariate analysis of accuracy scores indicated that the accuracy of SPD and the CPT improved significantly ($p < .02$ and $p < .03$, respectively). ST2 memory accuracy decreased to the extent that the analysis showed change, which approached the traditional level ($p < .06$). The overall data for accuracy on the ANAM tasks are summarized in Table 1. A significant change in response time was also found after field training, MANOVA, $F(7, 23) = 6.6$, $p < .001$. Univariate analysis indicated that SRT increased ($p < .03$) and that response time on SPD and one memory task (ST6) decreased ($p < .001$ and $p < .001$, respectively). All response time changes are presented in Table 1.

With two notable exceptions, these results are consistent with previous relevant findings (Callister et al., 1999; Elsmore et al., 1992; Slaven & Windle, 1999) that complex cognitive performance deterioration does not follow stressful exposure ubiquitously. Indeed, the accuracy of most tasks improved or was maintained after stressful exposure, and the ST6 was faster following stress. However, SRT was slower, and a trend toward less accurate performance on the least complex task, the ST2, was also found after the field exercise.

One explanation for the absence of complex cognitive task performance change after 1 week of stressful training coupled with the deterioration of SRT and a

TABLE 1
SERE ANAM Response Time and Accuracy

ANAM Scale	Accuracy					Response Time				
	Pre	*Post*	*Diff.*	*F*	*p* <	*Pre*	*Post*	*Diff.*	*F*	*p* <
SRT	100	100	0	—	—	224.3	234.7	10.4	5.16	.03*
CDS	96	96	0	<1	.623	1,014.0	1,007.3	–6.7	<1	.79
SPD	89.5	93.8	4.3	6.1	.015*	1,762.2	1,498.7	–263.5	17.8	.001*
LRS	88.1	86.5	–1.6	<1	.429	1,850.6	1,870.8	20.2	<1	.72
ST2	97.6	93.3	–4.3	3.8	.060	517.7	511.2	–6.4	<1	.73
ST6	92.2	93.2	1.0	<1	.497	758.1	674.9	–83.2	16.2	.001*
CPT	93.9	95.8	2.1	5.6	.029*	467.35	453.09	–14.26	4.04	.059

Note. SERE = Survival, Evasion, Resistance, and Escape; ANAM = Automated Neurological Assessment Metric; Diff. = difference; SRT = Simple Reaction Time; CDS = Code Substitution; SPD = Spatial Processing; LRS = Logical Reasoning; ST2 = two-item memory set; ST6 = six-item memory set; CPT = Continuous Performance.

*Denotes significant level of change.

Do not copy

low-demand memory task is that participants masked impaired cognitive functioning by increasing effort during the relatively brief assessment period, akin to the "end burst" effect in vigilance tasks. The absence of complex task change is consistent with reports that complex cognitive performance can be maintained during fatigue by increased effort during brief assessments (see Angus & Haslegrave, 1985). The decrements on the two least demanding tasks, SRT and the ST2, can be explained by a differential sensitivity of simple and complex tasks to effort. For example, Heuer, Spijkers, Kiesswetter, and Schmidtke (1998), among others, contended that increasing effort can maintain complex but not simple task performance. As increasing effort becomes more difficult as time on task increases (see Hancock & Desmond, 2001), the degree of performance decline during a trial would reflect this effort allocation required to maintain performance. Thus, if stress impairs information processing but the effect on complex tasks was masked by increased effort, performance decrements would be predicted to appear as time on task increases. Subsequent analysis was therefore performed on within-session performance change to explore this effort compensation hypothesis.

Analysis of Within-Session Cognitive Performance Changes

The within-session analysis compared accuracy and response time changes during the pre- and postfield training sessions. For each session, an average performance was calculated for the first- and second-half portions of trials, and the differences were subjected to analysis. There were 20 trials per subtasks, and because the first

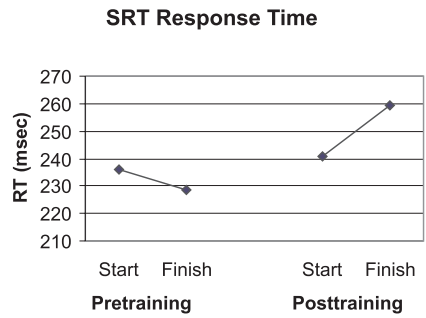


FIGURE 1 Within-session Simple Reaction Time (SRT) response time (RT) changes.

response is typically longer than responses during the remainder of the session, the mean of Responses 2 through 10 was compared with the mean of Responses 12 through 20. There were 64 trials in the CDS subtask, and Trials 2 to 21 were compared with Trials 45 to 64. Within-session changes during the pre- and posttraining sessions were then compared. The battery contained three types of tasks, and within-session performance changes are discussed separately for each type of task. SRT requires stimulus detection, and there is no manipulation or comparison of stimuli. In the previous analysis, mean posttraining reaction time was significantly slower than mean pretraining time ($p < .03$). The within-session analysis indicated that response time at the beginning of the pre-and posttraining sessions was similar; however, response time decreased during the pretraining session and increased during the posttraining session ($p < .05$). Figure 1 and Table 2 are illustrative confirmations of these effects.

Within-session changes of information manipulation tasks (CDS, LRS, and SPD) also differed during the pre- and posttraining sessions (Table 2). Although

Do Not Copy

TABLE 2
Within-Session Cognitive Performance Analysis

	Accuracy			*Response Time*		
Task	*Prechange*	*Postchange*	*p*	*Prechange*	*Postchange*	*p*
SRT				–7.31	18.2	.05*
CDS	–0.69	–1.10	.29	–115.4	107.0	.001*
SPD	0.90	–0.48	.001*	–25.50	15.84	.41
LRS	–0.17	–1.41	.008*	77.83	377.08	.006*
ST2	–0.28	0.62	.05*	15.32	37.18	.26
ST6	0.07	0.14	.75	–1.73	–85.41	.09
CPT	–1.0	–0.50	.36	–41.97	21.09	.008*

Note. SRT = Simple Reaction Time; CDS = Code Substitution; SPD = Spatial Processing; LRS = Logical Reasoning; ST2 = two-item memory set; ST6 = six-item memory set; CPT = Continuous Performance.

*Denotes significant level of change.

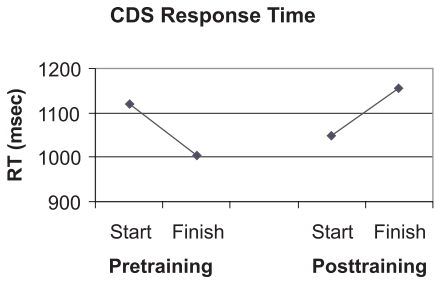


FIGURE 2 Within-session Code Substitution (CDS) response time (RT) changes during pre- and postfield training sessions.

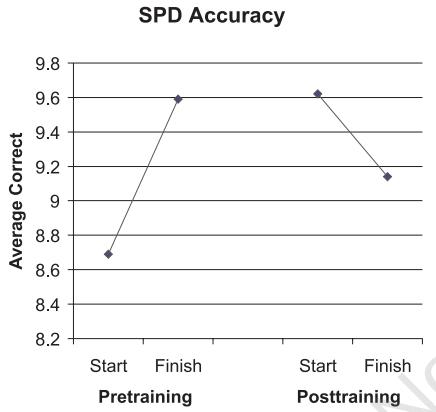


FIGURE 3 Within-session Spatial Processing (SPD) accuracy changes during pre- and postfield training sessions.

Do Not Copy

the specific change varied with the task, information management tasks were stable or improved during the pretraining assessment but then deteriorated during the posttraining session. CDS response time decreased during the pretraining trial but increased during the posttraining trial (Figure 2). SPD accuracy improved as the pretraining trial progressed and declined during the posttraining trial (Figure 3). LRS accuracy was relatively stable during the pretraining trials, but accuracy declined and response time increased as the posttraining trials progressed (Figure 4).

Significant differences between pre- and posttraining sessions were also noted for two of the memory tasks. ST2 accuracy was marginally worse ($p < .06$) during the posttraining session. In contrast to the within-trial pattern of information manipulation tasks, accuracy decreased as the pretraining trial progressed and increased in the later portions of the posttraining trial (Figure 5). The decreased accuracy in the posttraining session was therefore the result of poor initial performance during the posttraining session rather than change in performance during that session. The CPT showed response time decrease in the second session, which approached traditional significance level ($p < .06$). The within-session analysis indicated that response time decreased during the pretraining session but increased during posttraining trials ($p < .007$). Overall, the results of the within-session analysis suggest that previous studies did not find cognitive perfor-

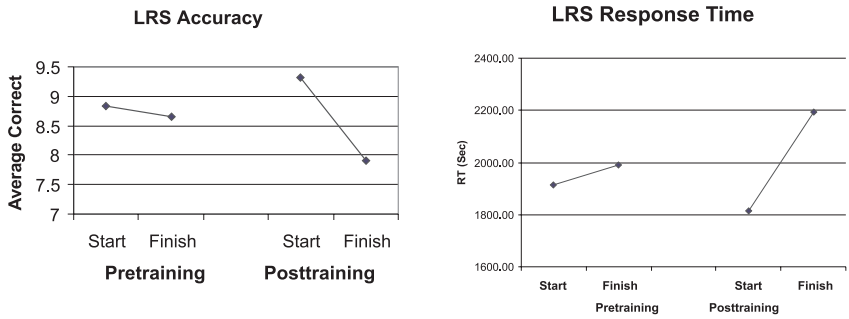


FIGURE 4 Within-session Logical Reasoning (LRS) accuracy and response time (RT) changes during pre- and postfield training sessions.

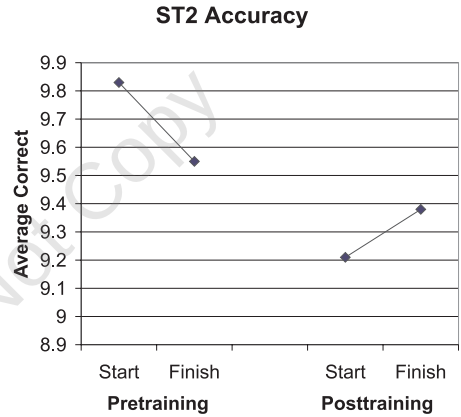


FIGURE 5 Within-session two-item memory set (ST2) accuracy changes during pre- and postfield training session.

Do Not Copy

mance decrements following stress because highly motivated participants were able to mask these complex task performance deficits by a transitory increase in effort during the relatively brief assessment interval. Because simpler performance tasks are relatively insensitive to the effects of effort, they do not show this change and thus show undisputed decrement effects.

Subjective State

For subjective state assessment, participants completed two paper-and-pencil measures, the POMS and the SOAP. Both measures indicated increased discomfort; however, there is no indication of serious psychological problems such as depression or tension/anxiety. Analysis of POMS data indicated that significant changes occurred during the week of survival training, MANOVA, $F(6, 29) = 32.44$, $p < .0001$. Univariate analysis indicated that Fatigue and Confusion scores increased and that the Vigor score decreased (all ps $< .001$). A summary of these changes is

TABLE 3
Profile of Mood States

Scale	*Pre*	*Post*	*Change*	*F*	*p* <
Tension	10.4	10.8	1.29	1.0	.321
Depression	7.5	6.6	1.37	1.43	.240
Anger	10.2	8.4	1.86	1.25	.272
Vigor	15.6	8.0	–7.83	50.4	.001*
Fatigue	45.4	67.5	19.31	179.6	.001*
Confusion	41.9	45.8	3.86	13.2	.001*

*Denotes significant level of change.

TABLE 4
SERE Students SOAP Changes

Scale	*Pre*	*Post*	*Change*	*F*	*p* <
Poor Concentration	20.0	26.3	5.97	21.96	.001*
Boredom	13.8	21.0	3.82	5.81	.022*
Slowed Reactions	15.6	28.6	12.47	85.85	.001*
Anxiety	18.59	19.91	1.32	1.08	.306
Depression	14.6	14.1	0.15	<1	.874
Irritability	18.7	19.8	3.06	4.72	.037*
Fatigue/Low Energy	16.6	37.2	21.82	219.82	.001*
Poor Sleep	21.0	30.7	15.15	64.51	.001*
Work Frustration	18.3	31.4	13.71	84.48	.001*
Physical Discomfort	39.0	24.32	250.93		.001*

Note. SERE = Survival, Evasion, Resistance, and Escape; SOAP = Sustained Operations Assessment Profile.

*Denotes significant level of change.

Do Not Copy

given in Table 3. A significant change occurred in SOAP profiles, MANOVA, $F(10, 24) = 33.8$, $p < .0001$. All scales changed significantly with the exception of the Anxiety and Depression subscales. As documented in Table 4, Fatigue and Physical Discomfort scales exhibited the largest changes.

Relationship Between Mood and Cognitive Performance Change

The effect of 1 week of SERE training that included physical discomfort, sleep disruption, and psychological stress produced reports of high levels of subjective discomfort and selective cognitive performance change. The significantly longer SRT and the marginal accuracy decrement on the ST2 suggest that these tasks are immediately sensitive to stress effects. A regression analysis was conducted to deter-

mine whether changes in these two cognitive tasks were related to subjective state self-report changes. SRT, the single response time that increased significantly following stress, was regressed on the POMS scales that changed during the week—Fatigue, Confusion, and Vigor. Analysis indicated a significant relationship between SRT and the three POMS subscales, $F(3, 15) = 4.41$, $p < .02$, adjusted $R^2 = .362$. The standardized coefficients and significance of the three scales were as follows: Fatigue, $\beta = -0.866$, $p < .30$; Confusion, $\beta = 0.627$, $p < .02$; and Vigor, $\beta = -0.292$, $p < .20$. Thus, although increased fatigue is the most prominent subjective change, it was the Confusion scale that was directly associated with increased SRT, not the Fatigue scale.

Although the level of accuracy change did not reach traditional levels of significance, the ST2 accuracy declined ($p < .06$), whereas slightly improved performance would be expected. The ST2 was therefore also regressed on the three POMS scales. Analysis indicated a significant relationship between the ST2 accuracy and the three POMS subscales, $F(3, 15) = 4.71$, $p < .02$, adjusted $R^2 = .382$. The standardized coefficients and significance of the three scales were as follows: Fatigue, $\beta = 0.355$, $p < .43$; Confusion, $\beta = 1.255$, $p < .06$; and Vigor, $\beta = 1.723$, $p < .01$. Decreased vigor was associated with decreased accuracy, but consistent with the relationship between mood and SRT, fatigue was not related to accuracy. Although the relationship between confusion and memory approaches significance, it is surprising to find that increased subjective confusion is associated with increased accuracy on the memory task.

Although performance changes are typically associated with fatigue (see Hancock & Desmond, 2001), the Fatigue scale was not related to presently recorded performance changes. However, decreased vigor and increased confusion were related to performance changes. This is consistent with some of our previous work that has indicated that changes in the Vigor scale of the POMS are most closely related to demanding physical performance (Bricker & Harris, 1999). It would appear that although people who have been exposed to an extended period of stress report that they are tired, concomitant reports of decreased vigor and thinking difficulty are better predictors of cognitive performance deterioration.

DISCUSSION

Military personnel perform physically and psychologically demanding tasks for extended periods during sustained and continuous operations, and the stress created under those conditions is expected to impair information processing (Wickens, 1996). However, several studies have surprisingly failed to detect significant acute cognitive performance deterioration in stressful military settings

Do not copy

(Callister et al., 1999; Elsmore et al., 1992; Slaven & Windle, 1999). This study examined cognitive performance change following a highly stressful military training exercise. Analyses provided an overall picture of cognitive performance variation that allowed testing of a number of subsidiary hypotheses. Subjective reports confirmed that the participants were physically uncomfortable after training and that fatigue increased and vigor decreased during the week of field training. In addition, participants reported difficulty concentrating in the posttraining session, a finding consistent with impaired information processing. In contrast to earlier studies, cognitive changes were found after 1 week of intense training. SRT was slower following field training. However, average complex task performance exhibited little change, which is consistent with earlier findings. However, when within-trial performance was examined to determine whether participants were maintaining poststress performance by exerting high levels of effort during the relatively brief assessment interval, it was found that transient levels of high performance could be maintained but that performance could not be sustained following chronic stress. The complex cognitive task performance deterioration within trials, which supports this interpretation, is consistent with the notion that performance at the beginning of the postfield assessment was at prestress levels as the result of increased effort (Angus & Haslegrave, 1985). Deterioration within the trial was thus the result of participants' decreasing ability to maintain effort with time on task (see Hancock & Desmond, 2001). Increased SRT would be predicted by Heuer et al. (1998), who contend that complex, but not simple, tasks can be maintained by effort. The within-trial analysis suggests that field training decreases participants' ability to continue to mobilize the resources required to perform complex tasks. Participants' ability to maintain performance levels appears to be quite limited as the number of stimulus presentations in each trial varied only between 20 and 64 and the time to complete a trial varied from 30 to 60 sec. The absence of cognitive deterioration in previous studies appears to be the result of assessment and data analysis procedures that masked complex cognitive task deficits by allowing transient effort to temporarily sustain performance level. In addition, assessments designed to detect changes in complex task performance have typically not included SRT, a task in which deficits are more difficult to mask. SRT may therefore provide an easily administered estimate of the effect of stress on complex cognitive performance.

Subjective reports are one tool to measure psychological state, but there is a reasonable concern that self-reports can be easily distorted and that they may not accurately convey psychological changes. In this study, subjective reports of vigor and ability to concentrate were related to cognitive performance variation. An unexpected finding was that although the dominant subjective change was increased fatigue, the fatigue changes themselves were not related to performance deterioration. These findings are consistent with reports that significant physical performance is related to vigor rather than fatigue (Bricker & Harris, 1999). The results of our study have implications for operators, equipment designers, and fitness for

duty testing. After periods of extended stress, military personnel can be expected to have difficulty operating equipment requiring complex information-processing capacities, and as the change appears to be the result of a decreased capacity to exert effort, their difficulty should increase with the length of the task, the number of tasks, or the complexity of the task. When a brief assessment is necessary, increased SRT appears to be the best test, but when sufficient time is available to assess complex task performance, sufficient prestress training must be provided to ensure that performance has stabilized. Thus, trials must be of sufficient length to avoid the masking of deficits by increased effort, and within-trial changes should be assessed to determine whether performance deterioration is an ongoing process. The results of our study are consistent with the model that stress decreases resource reserves and that as time on task increases, the resources available to process information begin to decline rapidly (Hancock & Warm, 1989). Individuals prioritize tasks and allocate available resources to high-priority tasks when such resources become limited (Hockey, Wastell, & Sauer, 1998). Therefore, performance during the initial portion of posttraining sessions showed minimum deficits. However, as time on task increased, the effect of stress becomes apparent as less effort can be mobilized to perform tasks. Application of these findings is important for improved military functioning in all phases and theaters of operation.

Do Not Copy

ACKNOWLEDGMENTS

This research was supported by Office of Naval Research Grant N00014–99–1–1034 to P. A. Hancock. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Office of Naval Research. We thank Susan Chipman, the Office of Naval Research (ONR) Technical Monitor. We are grateful to the personnel of the SERE School, Brunswick, Maine, for all their cooperation and assistance. In particular, we thank CDR J. M. Doherty for his help and direction.

REFERENCES

Angus, R. G., & Haslegrave, R. J. (1985). Effects of sleep loss on sustained cognitive performance during a command and control simulations. *Behavior Research Methods, Instruments and Computers, 17,* 55–67.

Belland, K. M., & Bissell, C. (1994). A subjective study of fatigue during Navy flight operations over southern Iraq: Operation Southern Watch. *Aviation, Space, & Environmental Medicine, 65,* 557–561.

Benedetto, J., Harris, W. C., & Goernert, P. N. (1995). Assessing gender differences and normative data on a cognitive performance measure. *Proceedings of the Human Factors and Ergonomics Society, 39,* 968.

Berkum, M. M. (1964). Performance decrement under psychological stress. *Human Factors, 6,* 21–30.

Bricker, T. A., & Harris, W. C. (1999). Managing workload with fatigue self-reports. *Proceedings of the Human Factors and Ergonomics Society, 43,* 1401.

Callister, J. D., Percival, G. L., & Retzlaff, P. D. (1999). Stress and fatigue during aircrew survival training. *Proceedings of the International Symposium on Aviation Psychology, 10,* 843–847.

Elsmore, T. F., Naitoh, P., & Linnville, S. (1992). *Performance assessment in sustained operations* (Rep. No. 92–30). San Diego, CA: Naval Health Research Center.

Englund, C. E., Reeves, D. L., Shingledecker, C. A., Thorne, D. R., Wilson, K. P., & Hegge, F. W. (1987). *Unified Tri-Service Cognitive Performance Assessment Battery (UTC–PAB): 1. Design and specification of the battery* (Rep. No. 87–10). San Diego, CA: Naval Health Research Center.

Hancock, P. A. (1986). Stress and adaptability. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 243–251). Dordrecht, The Netherlands: Martinus Nijhoff.

Hancock, P. A., & Desmond, P. (Eds.). (2001). *Stress, workload and fatigue*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hancock, P. A., & Hoffman, J. E. (1997). Stress and cognitive workload. In *Tactical display for soldiers: Human factors considerations*. Washington, DC: National Academy Press.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519–537.

Heuer, H., Spijkers, W., Kiesswetter, E., & Schmidtke, V. (1998). Effects of sleep loss, time of day, and extended mental work on implicit and explicit learning. *Journal of Experimental Psychology: Applied, 4,* 139–162.

Hockey, G. R. J., Wastell, D. G., & Sauer, J. (1998). Effects of sleep deprivation and user interface on complex performance: A multi level analysis of compensatory control. *Human Factors, 40,* 233–253.

Mareth, T. R., & Brooker, A. E. (1982). Combat stress reaction: A concept in evolution. *Military Medicine, 150,* 186–190.

Marshall, S. L. A. (1947). *Men against fire*. New York: Morrow.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). *Manual for the Profile of Mood States* (Rev. ed.). San Diego, CA: EdITS.

Orasanu, J. M., & Backer, P. (1996). Stress and military performance. In J. Driskell & E. Salas (Eds.), *Performance under stress* (pp. 89–117). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Reeves, D. L., Winter, K. P., LaCour, S. J., Raynsfor, M., Vogel, K., & Grissett, J. D. (1991). *The UTC–PAB/AGARD stress battery: User's manual and system documentation* (Special Rep. No. 91–3). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Retzlaff, P. D., King, R. E., Marsh, R. W., & French, J. (1997). *The development of the Sustained Operations Assessment Profile (SOAP)* (Rep. No. AL/AO–TR–1997–094). Brooks Air Force Base, TX: Armstrong Laboratory.

Slaven, G. M., & Windle, C. M. (1999). Cognitive performance over 7 days in a distressed submarine. *Aviation, Space, and Environmental Medicine, 70,* 604–608.

Sternberg, S. (1969). Mental processes revealed by reaction time experiments. *American Scientist, 57,* 421–457.

Villoldo, A., & Tarno, R. L. (1984). *Measuring the performance of EOD equipment and operators under stress*. Indian Head, MD: Naval Explosive and Ordinance Disposal Technology Center. (DTIC No. AD–B083 850)

Wickens, C. D. (1996). Designing for stress. In J. E. Driskell & E. Salas (Eds), *Stress and human performance* (pp. 279–295). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wickens, C. D., & Flach, J. (1988). Human information processing. In E. Wiener & D. Nagel (Eds.), *Human factors in aviation* (pp. 111–155). New York: Academic.

Do Not Copy

Ross, J.M., Szalma, J.L., Oron-Gilad, T., Hancock, P.A. (2005) A Meta-Analytic Review of the Effects of Temperature on Performance *Abstract American Psychological Association Division 21/Division 19 and Human Factors and Ergonomics Society Potomac Chapter Annual Symposium on Applied Experimental Research*.

A Meta-analytic Review of the Effects of Temperature on Performance

J.M. Ross, J.L. Szalma, T. Oron-Gilad, & P.A. Hancock
University of Central Florida

Accurate modeling of the effects of stress on performance is critical for maximizing the effectiveness of simulation and training. The U.S. Army Research Laboratory (ARL) has developed several tools and models such as Improved Performance Research Integration Tool (IMPRINT) that represent system performance as a function of human performance. However, there are gaps in the research literature on the stress-performance relation that constrain the effectiveness of IMPRINT and application of stress theory in general. This study is one in a series of meta-analyses aimed at quantifying stress effects in order to enhance the estimates used by IMPRINT to assess performance degradation under stress. In the current study we examined the effect of temperature on performance in both hot and cold environments. Preliminary results of the 'bare bones' meta-analysis using a random effects model (Hunter & Schmidt, 2004) indicated an effect size of d= -.46, with the effect size for the effects of cold (d= -.50) being somewhat stronger than the effect of hot temperatures (d= -.44), but variance estimates indicated that these effects were not homogeneous. Primary moderators to be considered are task categories based on the IMPRINT taxons, the type of performance measure used (e.g., accuracy vs. speed), and the range of temperatures employed. Results indicate that while temperature clearly impacts performance, these effects are moderated by task characteristics and dependent measures used.



Multi-Disciplinary University Research Initiative

MURI-OPUS

Operator Performance Under Stress

Perception of Morphed Objects: Generating psychophysical functions for the evolution of a shape

J.F. Morgan, T. Oron-Gilad, P.A. Hancock, &
J.L. Szalma

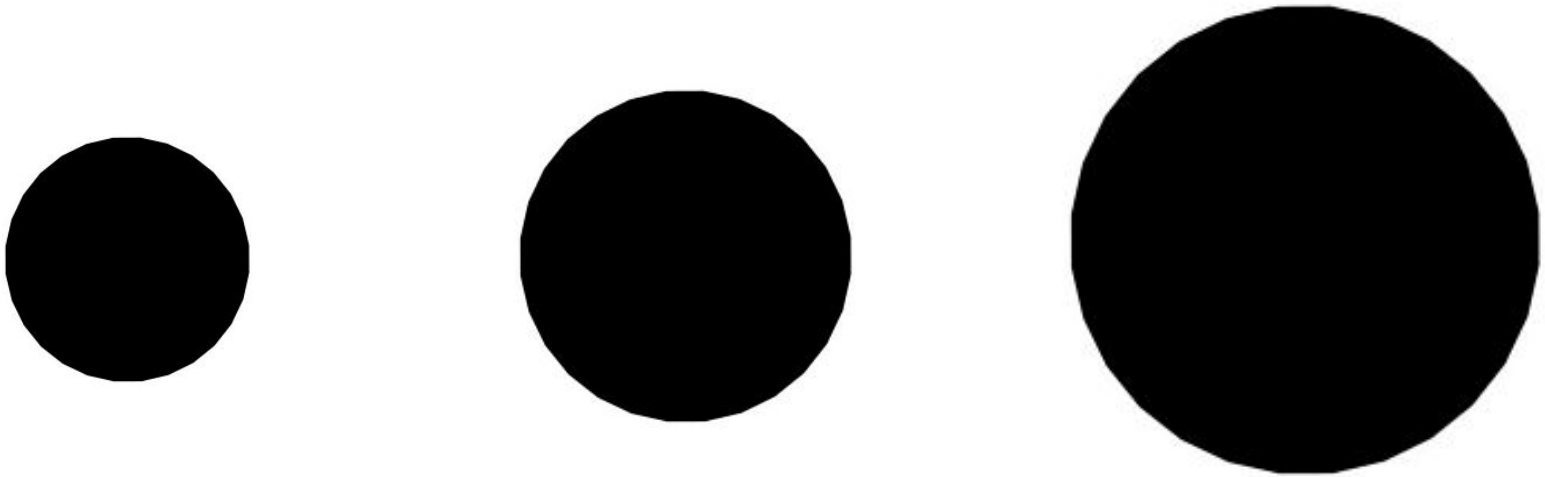
Background

- Serves as a baseline
- Friend/foe identification
- Information display

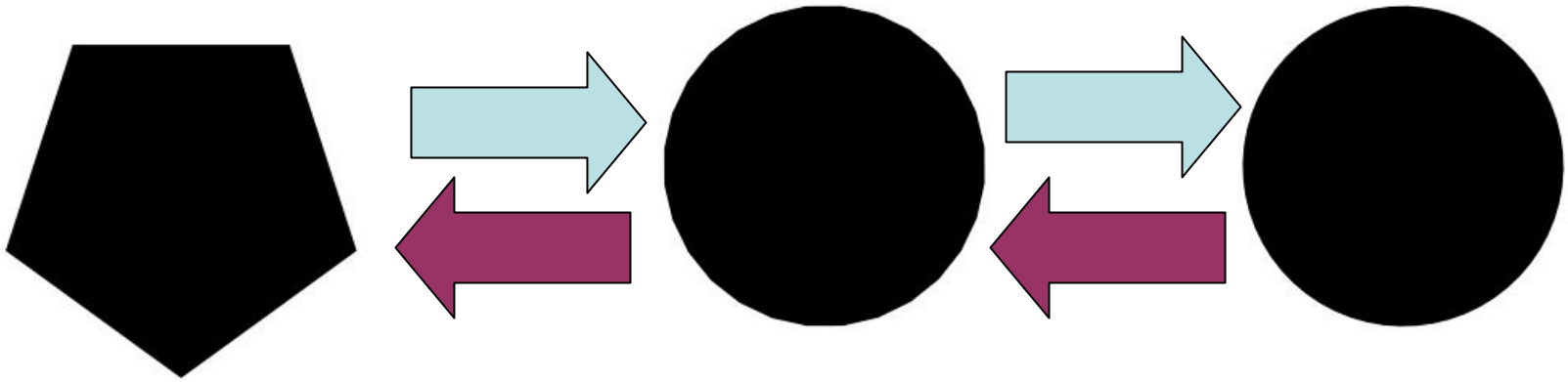
Method

- Psychophysical task using morphed stimuli
- Fechner's method of limits
- Instructions were written to lead participants towards more lenient responses
- 13 Participants
  - 9 females, 4 males
  - average age = 25.6

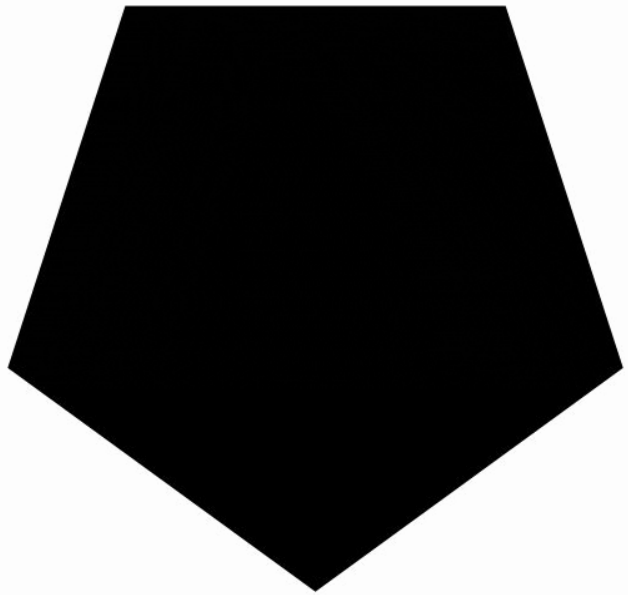
Stimulus Size



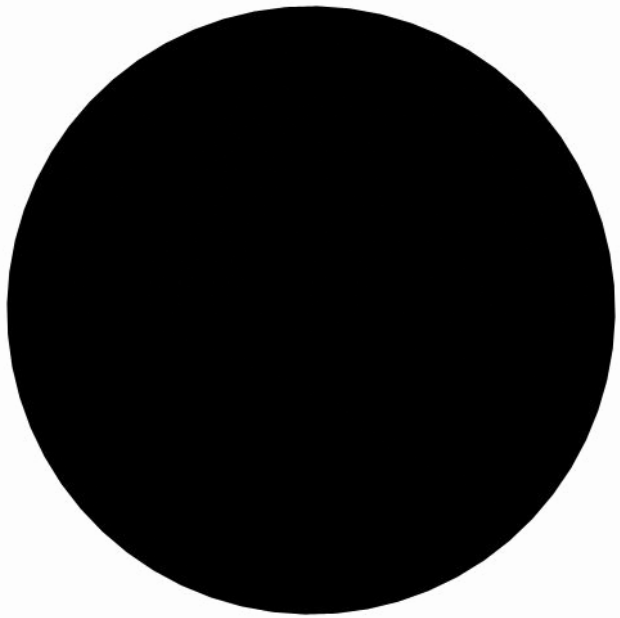
Direction



Task (Forward)



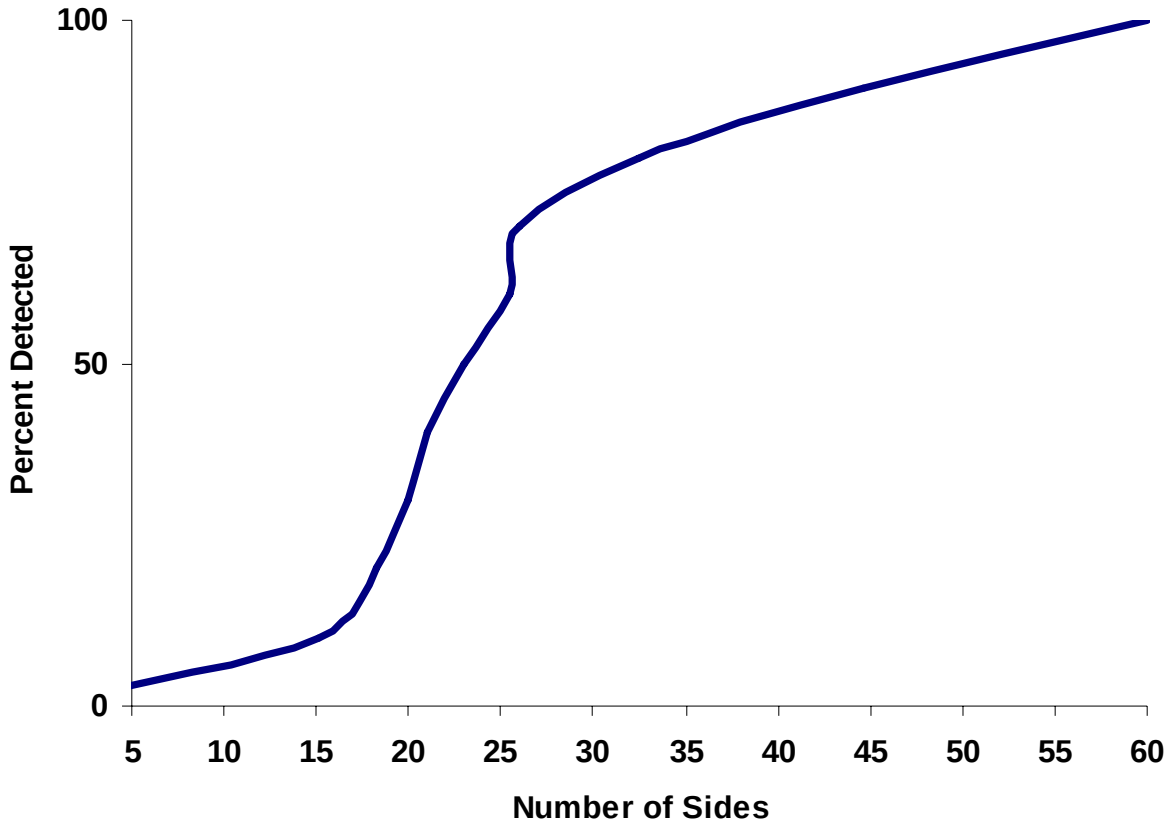
Task (Backwards)



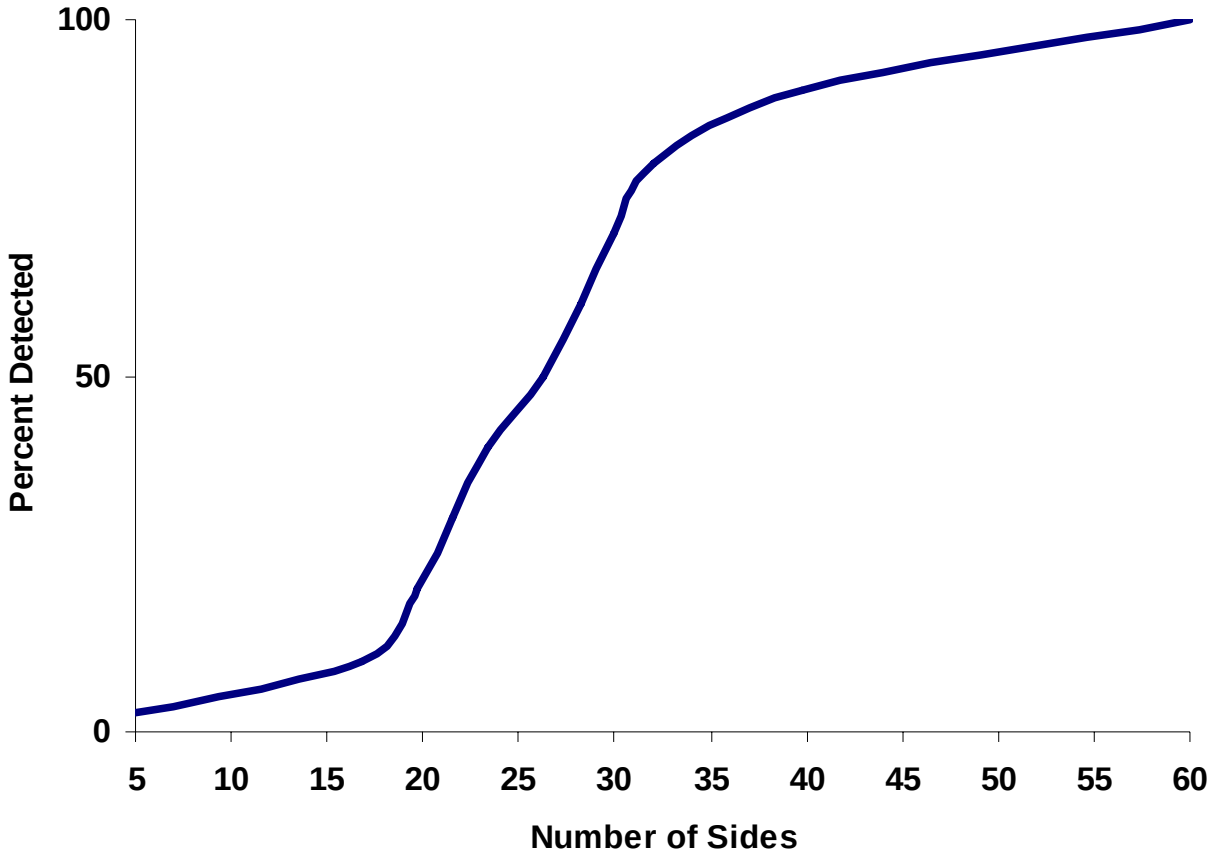
Results

- Instructions
  - When you "*perceive the shape becoming a…*"
  - Influenced participants towards more lenient behavior in the direction condition

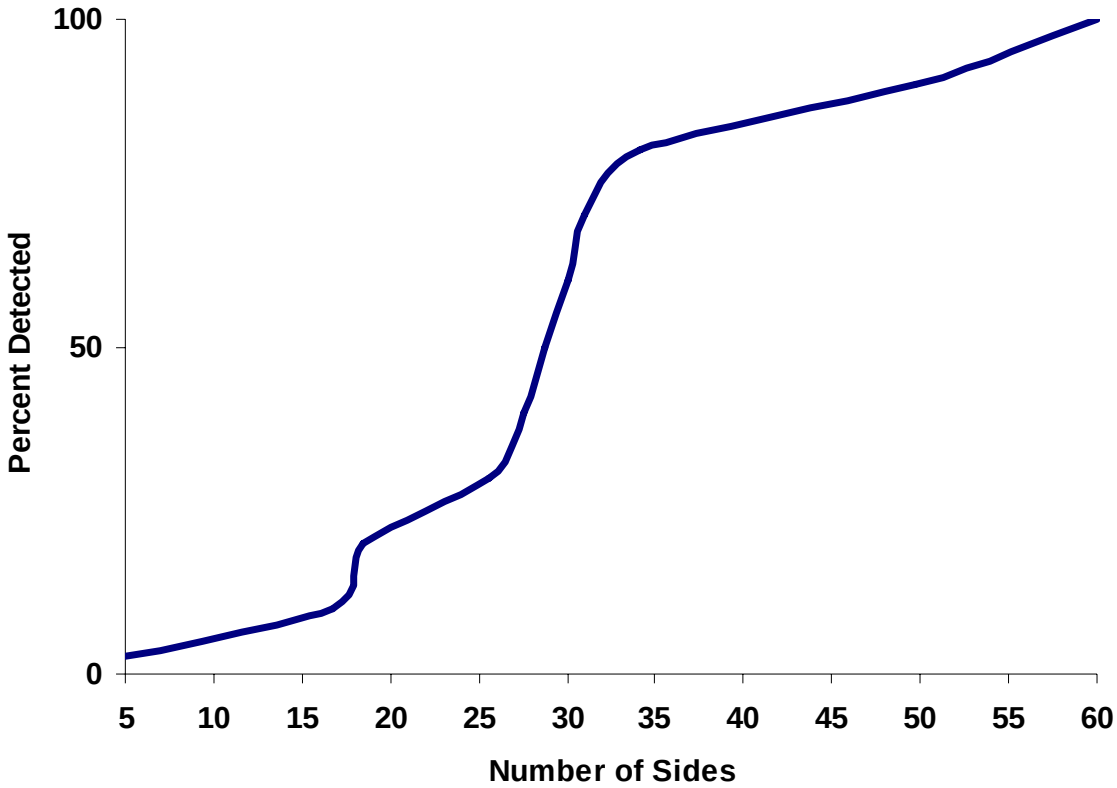
Size - Small


100
50
0
Percent Detected
5
10
15
20
25
30
35
40
45
50
55
60
Number of Sides

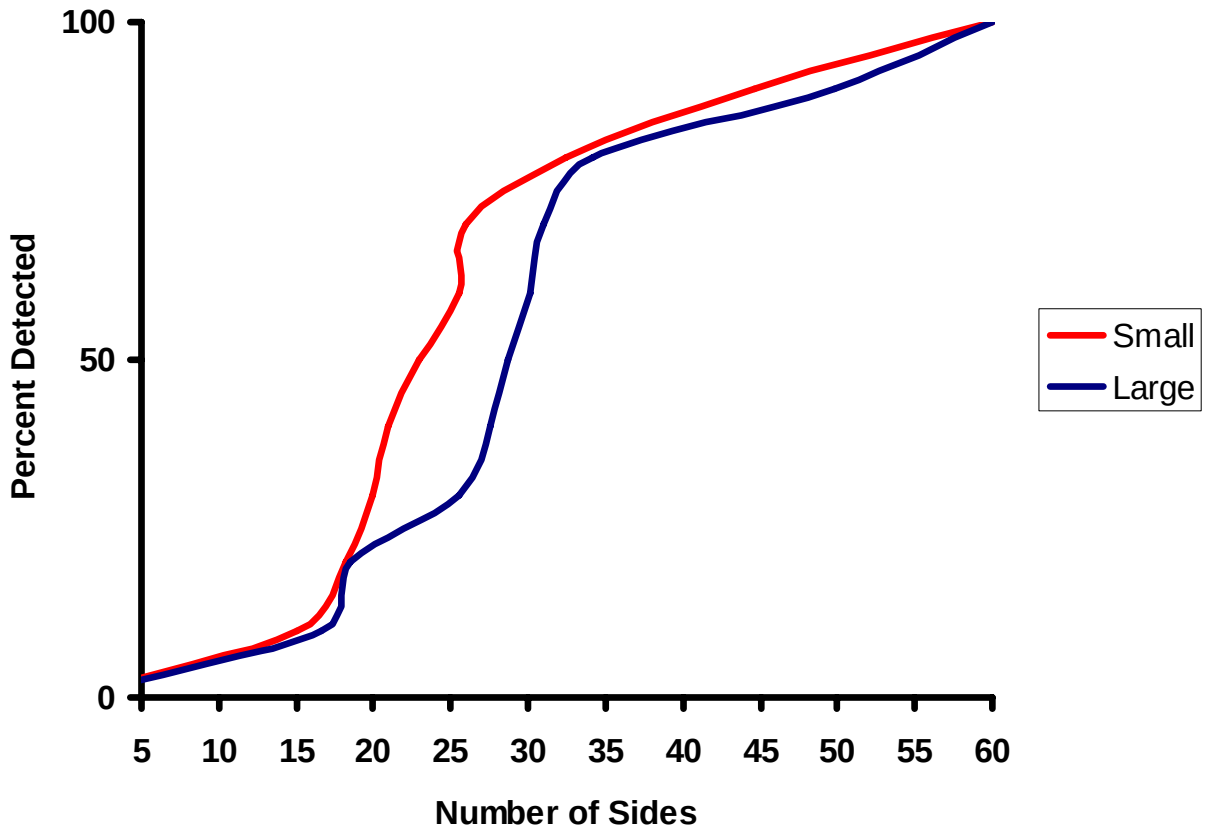
Size - Medium


100
50
0
Percent Detected
5
10
15
20
25
30
35
40
45
50
55
60
Number of Sides

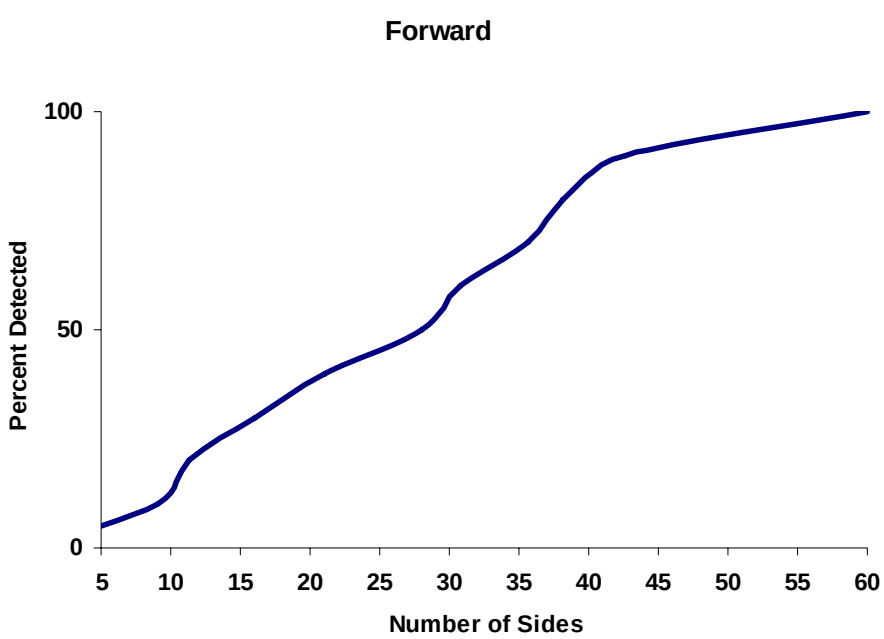
Size - Large

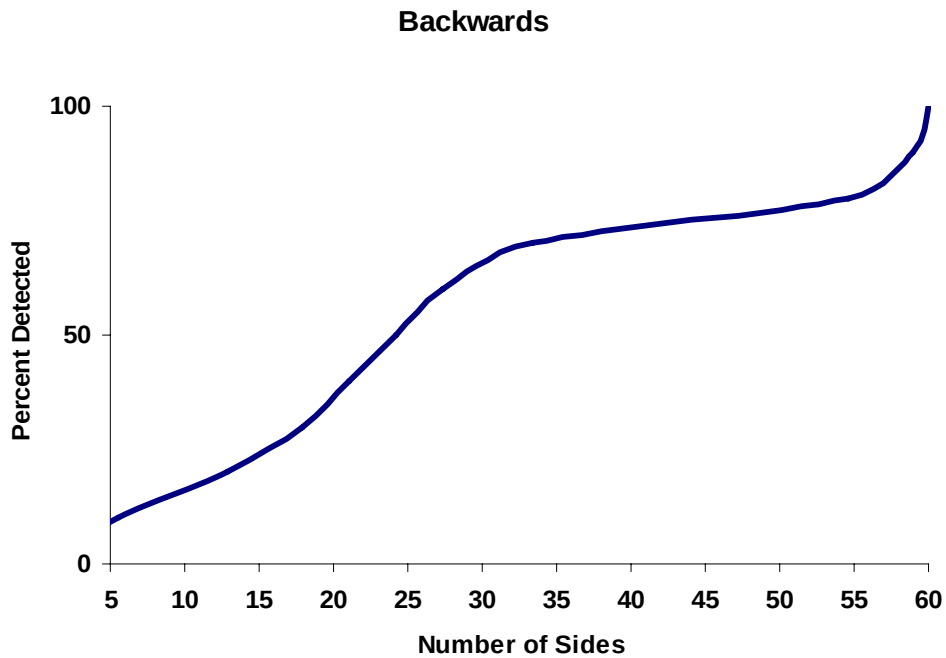

100
50
0
Percent Detected
5
10
15
20
25
30
35
40
45
50
55
60
Number of Sides

Size – Small & Large

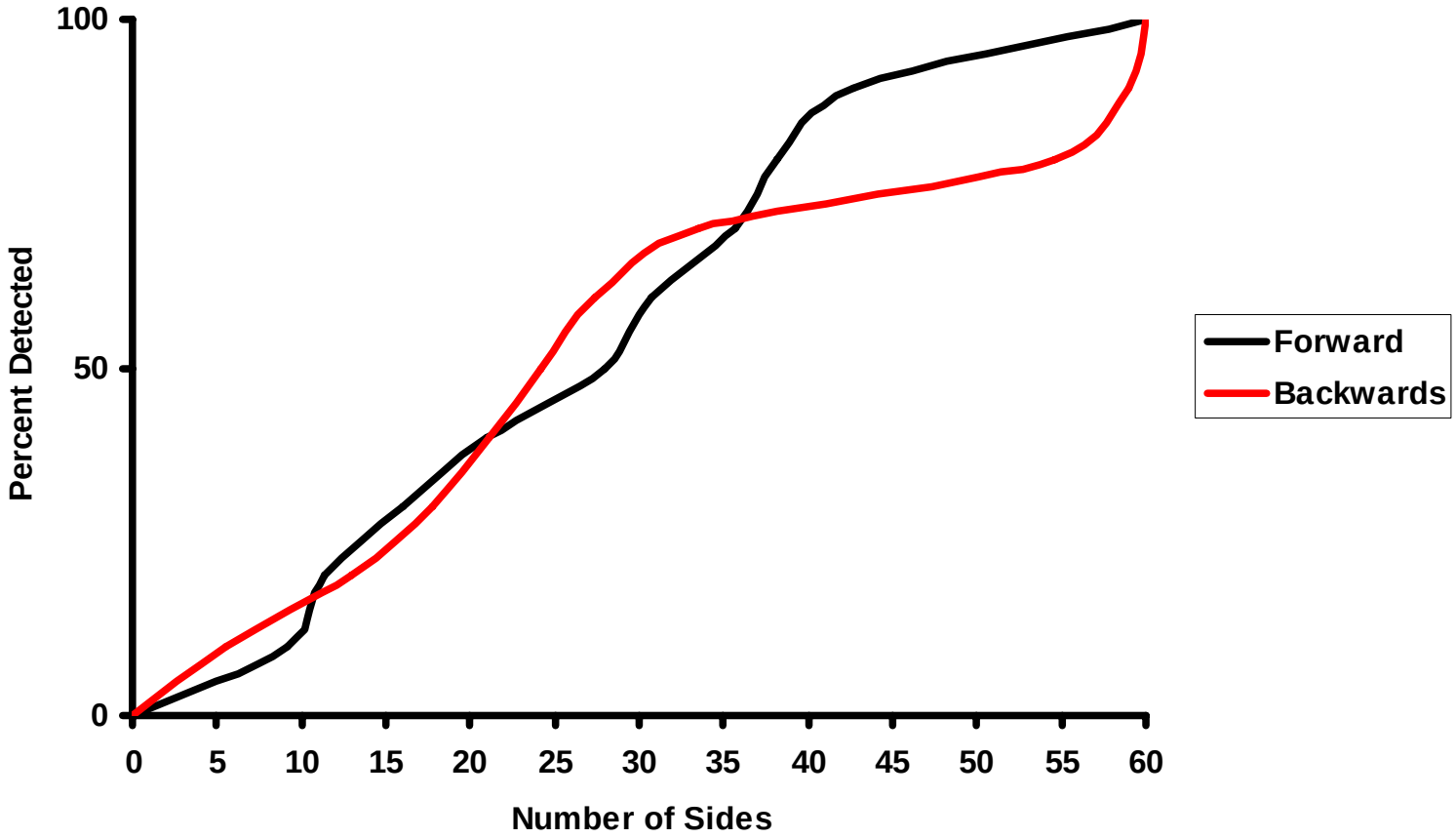


Direction


Forward
Percent Detected
100
50
0
5
10
15
20
25
30
35
40
45
50
55
60
Number of Sides


Backwards
Percent Detected
100
50
0
5
10
15
20
25
30
35
40
45
50
55
60
Number of Sides

Direction


100
50
0
Percent Detected
0
5
10
15
20
25
30
35
40
45
50
55
60
Number of Sides
Forward
Backwards



M1A1



T-55



75% M1A1,
25% T-55



25% M1A1,
75% T-55



50% M1A1,
50% T-55





M1A1



T-55



75% M1A1,
25% T-55

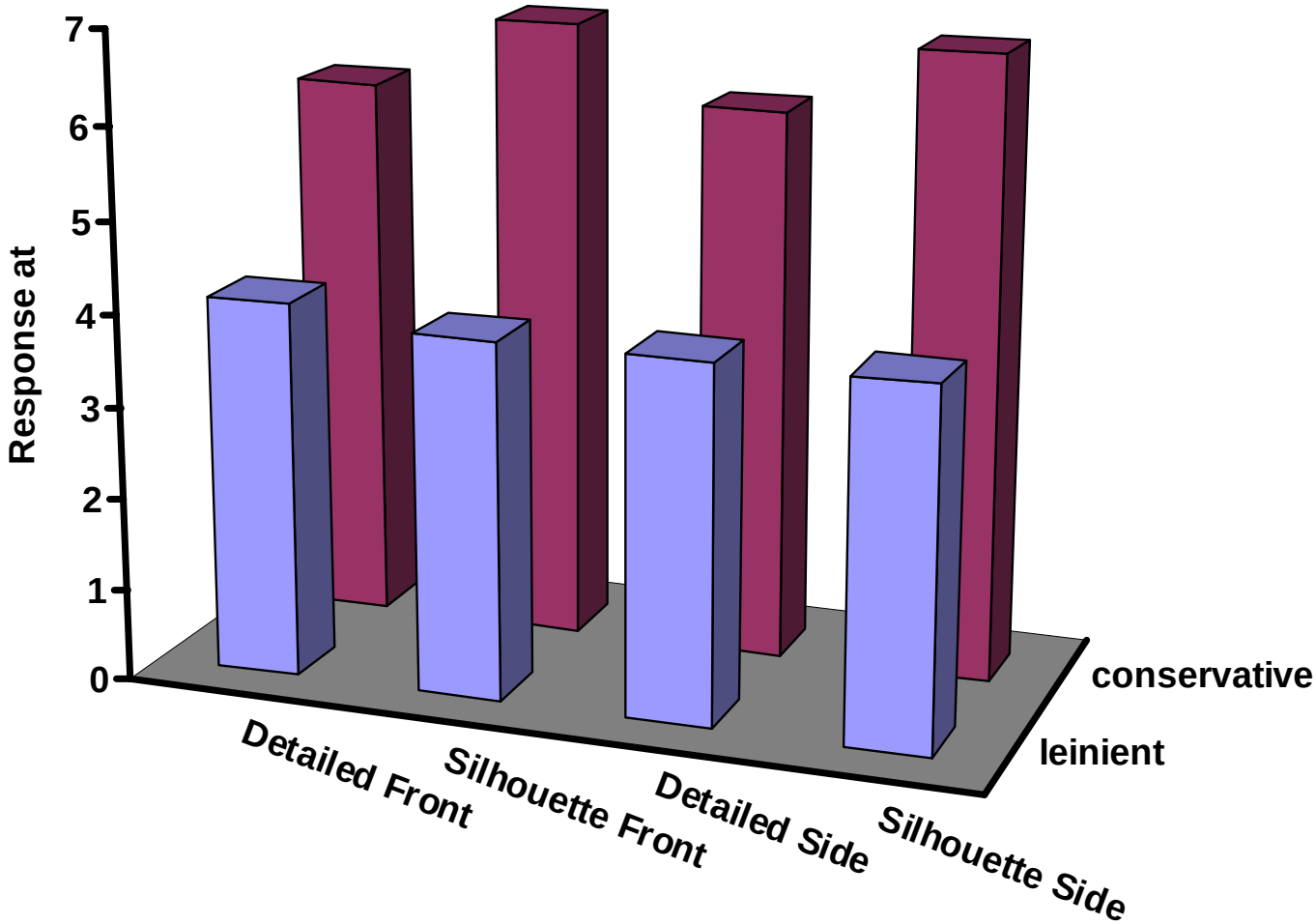


25% M1A1,
75% T-55



50% M1A1,
50% T-55




Response at
7
6
5
4
3
2
1
0
Detailed Front
Silhouette Front
Detailed Side
Silhouette Side
conservative
leinient

Conclusion

- Scalability
  - Size versus information
- Resolution
  - Basic versus complex
  - Emergent features

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621.

Szalma, J.L. & Hancock, P.A. (2005) Stress, Workload, and the Detection of Threat *Abstract Bugging Bugs and Brains: Communication and Applications from Cellular internals to Mental Contents, A Workshop/Symposium sponsored by the U S Army Research Office.*

Stress, Workload, and the Detection of Threat

James L. Szalma and Peter A. Hancock

Our program of research aims to further the development of psychological theory of stress, workload, and human performance, and to articulate potential strategies for the mitigation of stress effects in combat environments. The theoretical framework guiding this research is the maximal adaptability model of Hancock and Warm (1989). Central tenets of this model include, 1) Contrary to the inverted-U hypothesis (e.g., Easterbrook, 1959), individuals' adaptive behavior is stable over a wide range of stress levels. 2) Failures in adaptation can decline rapidly and progress from cognitive/affective state through behavioral outputs to the final, physiological failure. 3) Tasks themselves pose the proximal stress faced by individuals.

A major source of workload and stress in modern warfare comes from the ambiguity and uncertainty regarding the level of threat. Soldiers must detect enemy forces embedded in civilian populations while maintaining a focus on primary mission goals. Addressing these issues requires development of more powerful measurement tools for assessing threat detection performance under high workload and stress. It also requires an understanding of the individual differences variables that influence performance, workload, and stress, and how those variables interact with contextual variables. It is anticipated that these efforts will lead to improvements in the utility and power of stress theory, and from there development of techniques for mitigating the negative effects of stress. In this presentation we will describe efforts toward improved measurement of performance using Fuzzy Signal Detection Theory (Parasuraman, Masalonis, & Hancock, 2000), as well as advances in our understanding of the effect of personality traits on performance, workload, and stress.

References

Easterbrook, J.A. (1959).The effect of emotion on cure utilization and the organization of behavior. *Psychological Review, 66,* 183-201.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519-537.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42,* 636-659.

Individual Differences in Workload, Stress, and Coping in Police Officers Engaged in Shooting Tasks

*J.L. Szalma, T. Oron-Gilad, and P.A. Hancock**
The Minds in Technology – Machines in Thought Laboratory
Department of Psychology and Institute for Simulation and Training
University of Central Florida
Orlando, FL 32816

Abstract

In this study we examined relations between the personality traits of police officers and performance, workload, stress, and coping on four shooting tasks that were part of police officers' regular training. Expected trait effects on performance were not observed, but results indicated that personality traits influence how officers respond to shooting tasks. However, these effects are multidimensional and depend in part on task characteristics. This study represents the only explorations to date of the effect of personality traits on workload and stress in a shooting task using police officers. Future research needs to examine task characteristics that influence these relations among officers.

Introduction

Prior research has examined personality traits of police officers and their relation to global measures of performance (e.g., Barrick, Mount, & Judge, 2001) and job stress (e.g., Brown & Campbell, 1994). However, these studies have focused on global measures of performance and stress. To date, no study has examined the workload, stress, and coping strategies associated with shooting performance, a crucial skill among police officers. This study sought to explore these relations in four shooting tasks which were part of police officers' regular training.

Methods

Participants
The shooting tasks were part of required training for all police officers in the participating police department. Of the 91 officers in the training exercise, 71 (10 women and 61 men) volunteered for participation in the study. As a result of missing data, analyses for the current study utilized 58-65 participants, depending on the analysis. Mean age was 37 years and mean years of police experience was 11.

Task Design
The night shooting exercise included four different shooting tasks described below.

Warm-up Task (WT) Each officer was required to fire at a standard paper silhouette target. All rounds were fired in very low ambient light conditions, and performance on this task was evaluated by the percentage of hits (i.e., shot placement consistent with a lethal or incapacitating wound) out of the shots fired.

Flashlight Task (FT; see Figure 1*).* Officers were required to fire their weapons while holding a flashlight and rotating the body position to face 90 degrees or 180 degrees away from the target. Performance on this task was evaluated by the percentage of hits out of the shots fired.

The 'Barrel' Task (BT). Each officer was required to sequentially shoot at five targets, arrayed horizontally and equally distributed over approximately 90 feet. After engaging each target, the officer ran to the next shooting location to engage the next target. On targets two and four, the officer took cover behind a plastic barrel prior to engaging the target. Performance was evaluated by the percentage of targets hit.

The 'Metal' Task (MT). At the beginning of this task the officer was standing with his/her back against a wooden wall in a grassy area of the range. When cued the officer had to draw both flashlight and gun and turn 180 degrees and look through an opening in the center of the wall. Upon the command of the instructor the officer was required to fire twice at each of two targets of a specified color among three pairs metal targets. One pair of targets was white, one pair was red, and one was blue. Following this portion of the task the officer moved around the wall and engaged two sets of four targets. Each set of targets had a white target, black target, yellow target, and red target. Upon command, each officer fired twice at the targets of the color called out by the instructor. Performance on this task was evaluated by the percentage of hits out of the total shots fired.

Questionnaire Measures
Traits were measured using a set of adjective markers for the Five Factor Model (Goldberg, 1992), and yielded measures for Extraversion, Emotional Stability (Neuroticism), Conscientiousness, Agreeableness, and Intellect (Openness to Experience). Perceived workload was measured using a revised version of the NASA-TLX (Hart & Staveland, 1988), a well established measure of subjective workload that provides a global workload score as well as six subscale scores for Mental Demand, Physical Demand, Temporal Demand, perceived Performance, Effort, and Frustration. Stress state and coping were assessed using the Dundee Stress State Questionnaire (DSSQ), a multidimensional instrument to measure stress state and coping strategies (Matthews et al., 1999; Matthews & Campbell, 1998). The DSSQ measures three stress state dimensions: Distress, Task Engagement, and Worry. Note that these states were measured both prior to task performance and after the four tasks were completed. The coping scales measured are those of Task-focused, Emotion-focused, and Avoidant coping (see Matthews & Campbell, 1998) and are measured only after task performance.



Figure 1. Police Officers engaged in the 'Flashlight' Task.

Results

Regression analysis was used to evaluate the relation between trait measures and performance, workload, and stress measures. No significant effects of trait were observed in regard to performance across the four tasks. In regard to perceived workload, significant effects were observed for Conscientiousness, Emotional Stability, and Intellect, but the effects depended on TLX subscale and task. Thus, Conscientiousness predicted greater Temporal Demand for the FT task (see Figure 2), and Emotional Stability predicted increased Performance Workload for the MT (see Figure 3).The strongest effect on perceived workload was observed for Intellect, as shown in Table 1. In can be seen in the table that higher Intellect scores were associated with increased workload. However, these effects varied according to task. No significant effects on perceived workload were observed for the Extraversion or Agreeableness scales.

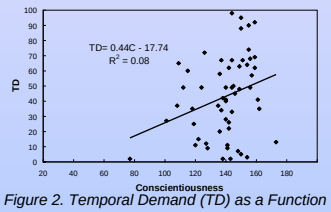


Figure 2. Temporal Demand (TD) as a Function of Conscientiousness (C): FT Task

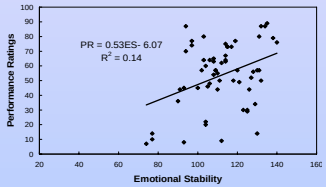


Figure 3. Performance Workload (PR) as a Function of Emotional Stability (ES): MT Task

Table 1. Regression of Perceived Workload Scales on Intellect (Openness)

Task	GW			MD			TD			P			E		
	b	SE_b	*β*	*b*	SE_b	*β*	*b*	SE_b	*β*	*b*	SE_b	*β*	*b*	SE_b	*β*
WT	ns	---	ns	ns	---	ns	ns	---	ns	.34	.16	.28	ns	---	ns
FT	.26	.10	.32	.32	.15	.27	.40	.16	.31	ns	---	ns	ns	---	ns
MT	.26	.11	.28	.39	.15	.32	ns	---	ns	ns	---	ns	.35	.13	.33

Note. GW=Global Workload; MD=Mental Demand; TD=Temporal Demand; P=Perceived Performance; E=Effort. ns=non-significant regression.

Higher pre-Task Engagement was predicted by higher scores on the Extraversion (b=.01, SE_b=.01, β=.35, *p*<.01) and Intellect (b=.01, SE_b=.004, β=.32, *p*<.05) scales. The other three traits did not significantly predict differences in pre-Task Engagement. However, with the exception of Agreeableness, the higher trait scores predicted decreased pre-task Distress among officers. Thus, higher levels of Extraversion (b=-.01, SE_b=.01, β=-.26, *p*<.05), Conscientiousness (b=-.01, SE_b=.01, β=-.26, *p*<.05), Emotional Stability (b=-.02, SE_b=.01, β=-.30, *p*<.05), and Intellect (b=-.01, SE_b=.01, β=-.27, *p*<.05) were associated with lower pre-task Distress. None of the five traits significantly predicted differences in pre- or post-task Worry.

Higher levels of post-Task Engagement were associated with higher levels of Intellect (see Figure 4) and Agreeableness (see Figure 5) In both cases, these effects were significant even after the pre-task state was accounted for in the model.

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, P.A. Hancock, *Principal Investigator*. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

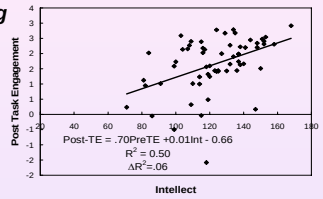


Figure 4. Post-task engagement (Post TE) as a function of Intellect (Int)

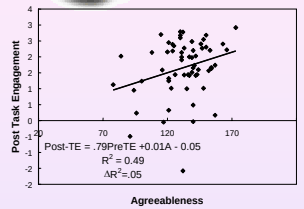


Figure 5. Post-task Engagement (Post TE) as a function of Agreeableness (A)

For Post-task Distress the only significant effect was for Emotional Stability (b=-.02, SE_b=.01, β=-.28, *p*<.05). However, these effects were not significant when the pre-task state was entered first into the model, indicating the effect of this trait on post-task Distress was mediated by pre-task state.

Regression analyses indicated that greater task-focused coping was predicted by higher Agreeableness scores (b=.09, SE_b=.04, β=.30, *p*<.05) or Intellect (see Figure 6). Similar analyses indicate that emotion-focused coping was more likely to be used for individuals who scored low on Emotional Stability (b=-.08, SE_b=.03, β=-.30, *p*<.05) and Intellect (b=-.05, SE_b=.02, β=-.26, *p*<.05). Avoidant coping was significantly predicted by Intellect (b=-.04, SE_b=.02, β=-.25, *p*<.05), such that higher levels of Intellect predicted decreased use of this coping strategy. No other effects involving traits or coping strategies were significant.

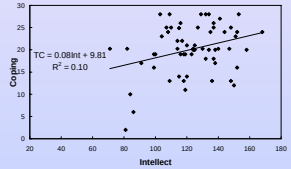


Figure 6. Task-focused coping (TC) as a function of Intellect (Int)

Discussion and Conclusion

➢Expected trait effects on performance were not observed, but traits were related significantly to perceived workload, stress, and coping. However, these results were not uniform across the four tasks.

➢Conscientiousness and Intellect were associated with higher perceived workload, as one might expect if such individuals are interested in the task or in performing well.

➢Emotional Stability was positively related to workload. It may be that those low on that trait experience less workload because they engage more in emotion-focused coping and therefore were less sensitive to task load.

➢DSSQ results revealed that officers' anticipation of task performance (pre-state) was more strongly influenced by traits than post-task state.

➢Although most studies of individual differences in personality and task performance have focused primarily on Extraversion and Neuroticism, the results of the current study indicate that Intellect may be a useful indicator of how well police officers cope with shooting tasks.

➢Further research needs to examine the degree to which these effects are task specific.

➢The tasks characteristics that influence these relations need to be identified.

References

Barrick, M.R., Mount, M.K., & Judge, T.A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Personality and Performance, 9*, 9-30.

Brown, J.M., & Campbell, E.A. (1994). *Stress and policing: Sources and strategies*. Chichester: Wiley.

Goldberg, L.R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4*, 26-42.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P.A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp.139-183). Amsterdam: North-Holland.

Matthews, G., & Campbell, S.E. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society 42*, 821-825.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state 'big three'? In: I. Mervielde, I.J. Deary, F. DeFruyt, and F. Ostendorf (Eds.), *Personality psychology in Europe (vol 7)* (pp. 335-350). Tilburg: Tilburg University Press.

Incorporating Individual Differences into the Adaptive Automation Paradigm

T. ORON-GILAD, J.L. SZALMA, J.E. THROPP & P.A. HANCOCK
Department of Psychology and Institute for Simulation and Training
University of Central Florida, Orlando, FL

Abstract

Adaptive Automation (AA) is the strategy which currently holds the greatest promise to achieve human-machine symbiosis. Individual differences are usually evaluated in the stage of operator selection but often ignored in the process of automation implementation. Our claim is that trait and state measures have a significant impact on how time and effort are allocated to required performance tasks. Thus, the aim of this work is to link trait reactivity with attentional indices that predict behavior and performance in computer mediated environments. How such an achieved symbiosis would affect the design of collective (human-human) organizations is considered briefly.

1. Introduction

Can human and automation work in a more compatible and harmonious manner? The answer is yes. However, to achieve such a goal, the engineers and designers of automation will need to know much more about the human operator. The first generations of automation were fixed in a static form of task allocation, in which there was a fixed designation for either human or machine task completion (Hancock & Chignell, 1989; Hancock & Scallen, 1996). Conversely, flexible allocation, in which task responsibility can be switched between parties as needed, suggested that adaptability would be a more effective approach. In adaptive systems, task allocation is dynamic and reflects changing demands posed by the task, the system, and the operator (see Figure 1). Figure 2 presents ten possible automation levels as proposed by Parasuraman, Sheridan and Wickens, 2000.

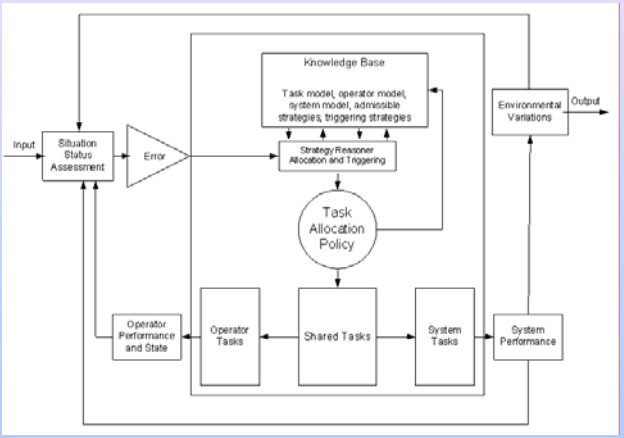


Figure 1: A preliminary framework for the comprehensive AFA architecture.

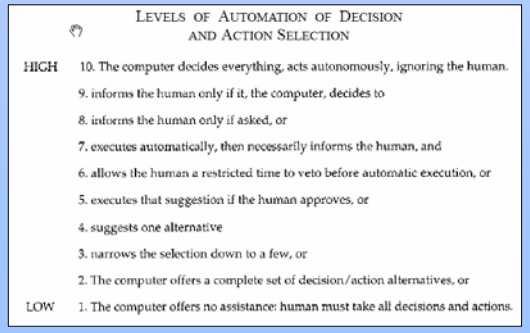


Figure 2: Level of automation of decision and action selection from Parasuraman, Sheridan and Wickens (2000).

2. Importance of System Individuation

Human operators may be more prone to succeed or fail in certain settings, yet this performance variability could be not only a factor of the system, but also a factor of the individual. An understanding of individual differences can be used to anticipate sources of human error, behavior, and preferences.

•Individual differences can determine **how well a job can be performed**. This would allow for increased productivity and more opportunities for different types of people to become users of the system.

•Individuation has often been circumvented by means of personnel selection, but this technique can not be used in situations where computerized systems are meant to be used by the public at large.

•**Training expenses can be reduced.** Rather than training all users to use the system in the same way, the system can accommodate different user coping behaviors.

•Advances in current technology allow for the accommodation of individual differences in different users. Through prediction and understanding of performance differences, systems can be developed which support the characteristics of a wider variety of individuals.

The information processing dimension of automation includes four levels: information acquisition, information analysis, decision selection, and action implementation (Parasuraman, Sheridan & Wickens, 2000); individual difference variables can affect each of these steps. Individual differences measures are being suggested to consideration within the individuation of adaptive automation paradigm. The information processing dimension of automation includes four levels: information acquisition, information analysis, decision selection, and action implementation (Parasuraman, Sheridan & Wickens, 2000); individual difference variables can affect each of these steps (as shown in Figure 3). Individual differences measures are being suggested to consideration within the individuation of the adaptive automation paradigm (as shown in Figure 4).

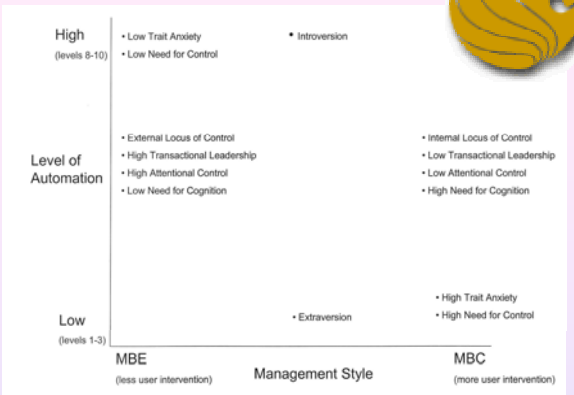


Figure 4:A matrix representation of relevant individual differences for level of automation and management style (management style represents the levels of human intervention in the process. MBE refers to management by exception where only in specific cases the operator would like to intervene in the automated process. MBC refers to management by consent, hence the operator approval is required at every stage of the automation, see Ruff, Narayanan, & Draper, 2002).

3. Conclusions

Adaptive automation that responds according to user individuation has many potential benefits, and is an area that warrants much additional investigation. Humans are unlikely to become obsolete in respect of many completely automated systems, such as a fully-automated commercial aircraft. The advantages of the human operator, such as complex pattern recognition and recall from past events to solve new problems, are difficult for a machine to replicate. The human will continue to be an integral part of automation, and so understanding and predicting their behavior, abilities, and limitations will continue to enhance safety and performance efficiency. A strong understanding of human individual differences can therefore serve as an important foundation for developing adaptive automation that optimally enacts human-machine interaction.

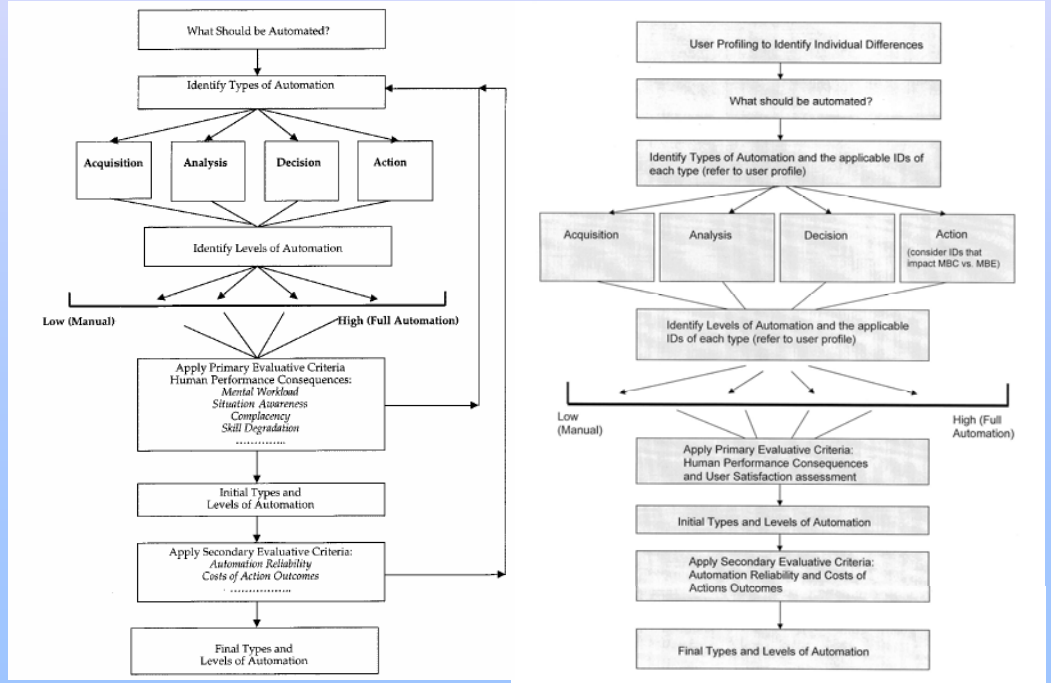


Figure 3:Figure 3(a) is a flow chart showing application of the models of types and levels of automation from Parasuraman, Sheridan and Wickens (2000). Figure 3(b) shows the individual profile assessment overlaid on this model.

Organizations per se, will not be bounded by the limits of space or time. To the extent of the database storage, one will be able to caucus with all who have previously contributed to a relevant state of knowledge and with fabricated avatars, virtual organizations become not simply feasible but preferable. We are now moving away from an era of generality into an era of specificity. We will soon no longer be able to talk of organizations as possessing simple common functions. They will become as individual, dynamic and fluid as their individual human constituents. In search of this future, the understanding of human individual differences remains a crucial enterprise.

Acknowledgments

This research was supported by a grant from the Human Research and Engineering Directorate (HRED) of the US Army Research Laboratory (ARL), P.A. Hancock, principal investigator. (Grant# DAAD19-01C-0065). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Michael Barns and Dr. Jessie Chen for providing administrative and technical direction for the Grant.

References

Hancock P.A. & Chignell,, M.H. (1989). Adaptive control in human-machine systems, *Human Factors Psychology,* P. A. Hancock, Ed. Amsterdam: North-Holland Publishing Co, pp. 305-345.

Hancock, P.A. & Scallen, S.F. (1996). The future of function allocation, *Ergonomics in Design, 4(4),* 24-29.

Parasuraman, R., Sheridan T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation, *IEEE Transactions on Systems, Man, and Cybernetics,* **30**, 286-297.

Ruff, H.A., Narayanan, S., & Draper, M.H. (2002). Human interaction with levels of automation and decision aid fidelity, *Presence*, *11*(4), 335-351.

Time, Emotion, and the Limits to Human Information Processing and the Organizations they Form

P.A. Hancock*, J.L. Szalma, & T. Oron-Gilad
The Minds in Technology – Machine in Thought Laboratory
Department of Psychology and Institute for Simulation and Training
University of Central Florida
Orlando, FL 32816

Abstract.

We propose a three-level, brain-based, mechanism to explain the fundamental phenomena of time. Coordination between these levels serves as a rate-limiting control on information processing capacity. Maximal human capacities are posited as a function of the amount of information to be processed, of best sensory processing speed, and of the Cognitive Mechanism of Knowledge Representation (CMKR). CMKR is constructed of both memory and generated, context specific, 'what-if' scenarios. Emotional states influence each of these processes primarily through sensory and cognitive mechanisms. Implications of this tri-level conception for the design of organizations are adumbrated.

1. Introduction

Individuals are only able to form themselves into organizations when they share some common values, assumptions, language, and information. The negation of such sharing is chaos, which in colloquial understanding is the antithesis of organization. One of the fundamental, shared properties of human reality is the concept of time. Our work here explores the concept of time and how human beings fabricate and process such fundamental information. The general structure of our model is shown in Figure 1. Hence, the trinity of time: time organization in the brain is composed of three levels as shown in Figure 2. The zero-order system and purpose is to provide gross synchronization with the environment while permitting a continuous '*stream of consciousness*' form of experience, being continuous in nature. We propose that there is a central oscillator that acts to generate its own independent, endogenous time signal (as Shown in Figure 3). This capacity has evolved in symbiosis with the temporal signals of the environment and so readily entrains to them. In the human, as with most advanced homeotherms, the signal is largely contingent upon the diurnal or light-dark cycle. Oscillation frequency depends upon brain temperature in the immediate area of the limbic system, which itself is heavily defended by homeostatic response. Manipulations that act to change core brain temperature influence oscillation frequency. These include overall body temperature itself (Hancock, 1984), drugs (Hicks et al, 1984), circadian phase (Aschoff, 1984), and stress and excessive task demand (Hancock & Weaver, 2005). It is the system which answers Hume's requirement for personal continuity. However, this system allows little in terms of fast response to environmental change or future planning about forthcoming challenges, which are each capacities of a higher order.

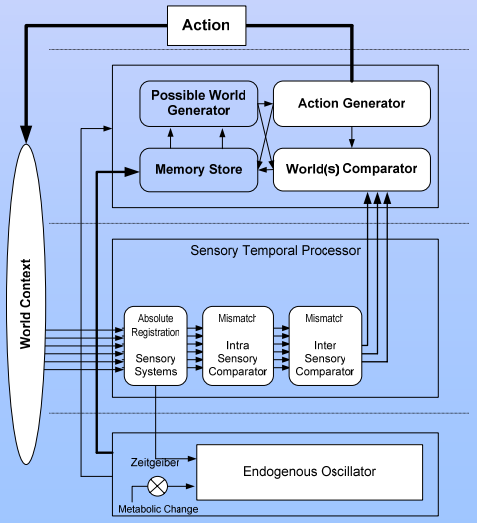


Figure 1. The Tri-Level Model of Temporal Processing in which process relationships are specified.

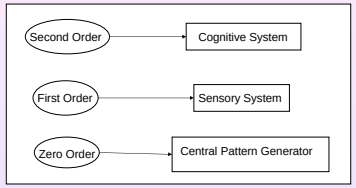


Figure 2. The Trinity of Time: The Organization of Time in the Brain

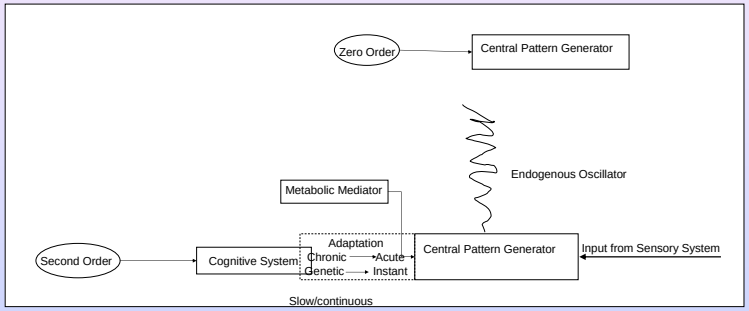


Figure 3. The Central Pattern Generator

2. Sensory Level Timing

If all temporal data had to be passed to a central generator for processing, such data would be limited by the fundamental frequency of that oscillator. However, timing in sensory systems provides much greater resolution than is possible with forced interaction with a central generator and thus we believe that there is a second 'action-based' timing capability erected upon the oscillator and partially integrated with this existing low-frequency system. A first process is intra-sensory comparison (shown in Figure 4). Accomplished on the millisecond level, intra-sensory comparison of the temporal structure of incoming stimuli allows orientation and localization in space to a staggering degree. Following the intra-sensory comparison, the inter-sensory comparison provides further comparative temporal information. These comparisons finesse the need to provide a unique 'now' in the brain. Remaining disparities are not a problem, in fact they also provide important temporal cues. Inter-sensory comparison is much slower than the intra-sensory comparison and so provides a longer time horizon. Finally, some degree of integration with continuity is necessary and we believe that this is the point at which synthesis with bottom-up temporal information from the central generator occurs. To summarize, the sensory timing system is represented as a threshold system in which sensory data is accumulated to an adaptive, sensory threshold. The sensory level system is for fast response and treats time as a series of 'windows of action opportunity' not a continuous stream. Thus, unlike the wave function that predominates at the lower level, the idea of a temporal 'quantum' is feasible at the sensory, first-order level.

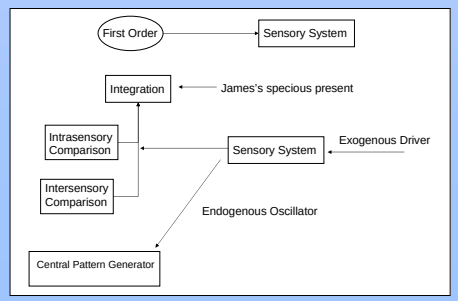


Figure 4. The sensory system



Twenty years ago, Peter Hancock takes refreshment at the first ODAM meeting!

> *"When a man sits with a pretty girl for an hour, it seems like a minute. But let him sit on a hot stove for a minute-and it's longer than any hour. That's relativity."*
>
> *Einstein*

3. The Cognitive Clock

If the sensory level of timing provided momentary registration of change, the data structures they provide about change in the environment are converted at the cognitive level to 'information.' This requires an integration of sensory-up transmission combined with cognitive-down selection to arrive at 'information' accumulation. Sub-served by memory, the purpose of the cognitive clock is for projection into the future. Being a second-order system, the 'cognitive clock' is concerned with three basic functions; founded upon previously accumulated information, it runs a series of 'what-if' scenarios based upon spatial and temporal horizons established by experience. We call this condition the 'normative zone.' In this zone, we believe that temporal information from all three levels is synchronized, the individual perceives no immediate threat, nor is there any immediate unfulfilled need. The time base generator is communicating a signal of around 20hz, providing a 'specious present' of approximately 50 milliseconds (See James, 1890; Vroon, 1973). Since the rate of information accumulation, founded upon the sensory communication, approximates the central-pulsed, time-base, there is no temporal dissonance of either time-in-passing, or time-in-memory.

At exactly the juncture that applied human performance researchers become interested, that is, when the pragmatic necessity to understand optimal information processing becomes vital, the synchronization between the tri-partite differentiations of temporal information begins to diverge. This can happen in a series of ways. First, as we have noted, the central generator can be affected by various influences that alters the fundamental time base. If the individual is not acting in an immediate goal directed way, this distortion will appear to be the contraction/dilation/distortion as reported in accounts such as those of Huxley (Huxley, 1954). These influences will certainly influence time-in-passing and, by implication, time-in-memory being recalled after the event. What has not previously been recognized to any great degree is the influence of such experiences on time-in-future. Our contention is that the prime function of the cognitive clock is to use past experience to run future scenarios (of different spatial and temporal horizons). Previous distortions can therefore act to affect the setting of that threshold.

4. Quantifying Human Information Processing and Optimal Performance

As we indicated earlier, the establishment of a numerical value for Human Information processing (HIP) might be of particular satisfaction and value to a pure scientist, but a single value would always be flawed since it could not account for novelty, context, experience, and any of a plethora of non-stationary factors. Quantification of human information processing is especially important in new, novel, uncertain, stressful or generally unusual circumstances. Our present persuasion is that the clocking mechanisms are linked but insufficiently vertically integrated (as are many brain systems). We believe that study directed to understand the temporal nature of processing activity will return a significant reward and will not simply follow previously tried and tested avenues of investigation.

Time is no simple linear record of events. Rather, it is multi-faceted panoply of opportunity with different evolutionary elements reflected in brain structure and function. Organizations have used time as a very constraining way to synchronize social interaction to a commonly imposed metric and have largely neglected the opportunities rendered when time is not viewed solely in this mechanistic fashion. By exploring the ways in which time is fashioned in the brain we hope to encourage similar exploration on a collective socio-organizational level. In this way we can introduce all the unexplored opportunities of time.

Acknowledgment

The views expressed in this work are those of the authors and do not necessarily reflect those of any named individual or agency. This work was supported under grant # NBCH 10030012 by the Defense Advanced Research Projects Agency (DARPA) via the Department of Interior. The Grant Monitors for this work were Dr. Dylan Schmorrow, Dr. Russ Shilling, and Dr. Jim Patrey.

References

Hancock, P.A., Szalma, J.L. and Oron-Gilad, T. (2005). Time, Emotion and the Limits to Human Information Processing. In Dennis McBride, and Dylan Schmorrow, (Eds.), Quantifying Human Information Processing

Complete list of references in proceedings paper.

Nayeem, R.V., Oron-Gilad, T., & Hancock, P.A. (2005) Operators' Time Perception under Stress *Paper Proceedings of the International Symposium on Aviation Psychology.*

Operators' Time Perception under Stress

R.V. Nayeem[1], T. Oron-Gilad[1&2], & P. A. Hancock[1&2]

Department of Psychology[1] and [2]Institute for Simulation and Training

University of Central Florida

ABSTRACT

Time perception is extremely important to the understanding, design and use of complex aviation systems. This experiment focused on differences in time estimation, flight performance, and monitoring tasks. In a between-subjects experiment, participants navigated through a flight scenario while monitoring a switch and listening to white noise at either 55dBA or 85dBA. Flight performance data and monitoring data were collected throughout the task. Participants also completed the NASA-TLX and the DSSQ-S. Statistical analyses showed that the noise condition did not significantly affect workload, monitoring abilities, task completion and subjective stress questionnaires for the dual task. However, the 85dBA condition significantly affected prospective time estimation. These results suggest that the dual task was not demanding enough and the stress was not adequate to push participants out of the comfort range and experience a performance decrement.

Key Words: Time perception, Workload, Stress, Noise

INTRODUCTION

The temporal domain is stress-sensitive in a similar manner to the spatial domain and comparable narrowing occurs, resulting in distortions of perceived time (Hancock & Weaver, 2005). Understanding time perception and time distortion under stress is of high importance for operating complex systems. Time perception affects how operators react to visual, auditory and tactile alarms. If time estimation is inaccurate due to stress, the alarm may go unattended for a critical length of time which exceeds the time available for solution. Time estimation is affected by the stress and mental effort that operators experience. If stress conditions are sufficiently high, such conditions will induce time distortion (Block, Zakay, & Hancock, 1999). This has been seen in Eastern Airlines Flight 401 in the Everglades where the crew got fixated on a landing light, and in the infamous John Denver fatal aircraft accident.

One possible solution to this problem is adaptive automation. Adaptive function allocation or dynamic automation is an approach in which control of tasks dynamically shifts between humans and

[1] Corresponding Address: Razia V. Nayeem, Department of Psychology, PO Box 161390, University of Central Florida, Orlando, FL 32816-1390 E-mail: rnayeem@ist.ucf.edu

machines, and is an alternative to traditional static allocation in which task control is assigned during system design and remains unchanged during operations. With adaptive automation, if operators are under too much stress to respond optimally to the aircraft, tasks will automatically get shifted to the automation rather than the operator. If automation states are switched too quickly, operators can become confused and decision making can be degraded, and allocation induced oscillations car occur (Hancock & Scallen, 1996). Also, when a trigger is operator-driven, the added workload and decision making can hinder performance (Hildebrandt & Harrison, 2002). Using dynamic function scheduling, an operator can prioritize re-allocation by assigning tasks both a temporal value, when the task should be re-allocated along the system timeline, and a qualitative value, what contribution the task makes to the system goal and the quality of the eventual solution. The Hancock and Warm (1989) model shows different zones in which a person can be when stressed.

The purpose of this study was to investigate differences on time estimation due to differences in the audible level of white noise, while performing a flight navigation task and a dual monitoring task. Stress was manipulated by task complexity, time constraints, and the addition of one of two levels of white noise: standard high frequency white noise at either 55dBA or 85dBA. This study attempted to investigate time distortion associated with stress and monitoring tasks in a basic flight navigation task. It was anticipated that participants would be pushed out of the comfort zone (Hancock & Warm, 1989) and would experience performance decrements to workload, monitoring abilities, subjective stress questionnaires, and prospective time estimation. However, given the relatively simple nature of the task and the qualifications for participation, it was expected that the noise would not affect task completion.

METHOD

Participants. 32 pilots, 28 male and 4 female students enrolled in psychology and human factors courses at Embry-Riddle Aeronautical University. The mean age was 21.19 (SD=3.61). Participants were required to have at least a private pilot's license. Total hours as PIC ranged from 40 to 700, with a mean of 133.1

(SD=134.1).Students who chose to participate in this research effort were compensated via extra credit. The participants all filled out personality questions as well. On a rating scale of 1-5, they uniformly rated themselves as able to pace themselves to get things done (M=4.0, SD=.6), able to handle stress (M=2.3, SD=.9), as moderately methodical people (M=2.4, SD=.8), and as having fast-paced lives (M=3.3, SD=1). About half of the participants responded that they waste a lot of time before settling down to work, and about half responded that they do not (M=3.9, SD=.7).

Apparatus and Tasks. The experimental setup consisted of two independent tasks generated in the AirBook® by Simigon F-16 flight simulator: a) the monitoring task and b) navigation task. The flight simulator was installed on a Dell laptop model # PP08L, with a 17" screen. Static white noise was emitted through headphones plugged into the computer speakers.

White Noise. There were two levels of white noise – 55dB and 85dB. The sound was delivered via Maxell headphones, model # HP/NC-II, and was measured before each participant via a Sper Scientific sound meter, model # 840029.

Flight Controls. Participants manipulated the flight environment using a GF yoke, model # G60503A.

Navigation Task. The navigation task objective was to take off from an airport, navigate to three waypoints, and land at the initial airport, as shown in Figure 1. A skyscraper marked each waypoint. After the participant crossed over the waypoint at 5000 feet MSL, the avionics of the aircraft changed to guide the participant to the next waypoint. When a participant crossed all three waypoints, the task was considered successfully complete. The waypoints were strategically placed throughout Arizona terrain. Two different navigational missions were used for the practice trial and the dual task trial so that participants would not become too familiar with any one mission. The navigation task difficulty did not vary from mission to mission. Additionally, at the end of the second mission, participants were asked to estimate how long they spent completing the dual task.

Monitoring Task. The monitoring task consisted of periodically checking the master switch in the aircraft console. The switch was set to turn to the "off" position 30 seconds after takeoff, and 30 seconds after the participant crossed the second waypoint. Switch states are shown in Figure 2. Reaction time and accuracy were recorded for the monitoring task. If the participant did not turn set the switch to the "off" position after the first change, the switch remained in this position for the entire flight. Similarly, if a participant did not set the switch to the "off" position after the second change, the switch remained in this position for the remainder of the flight. If participants did not cross the second waypoint at all, or higher than 5000 feet, the switch remained in the position it was prior to these events.

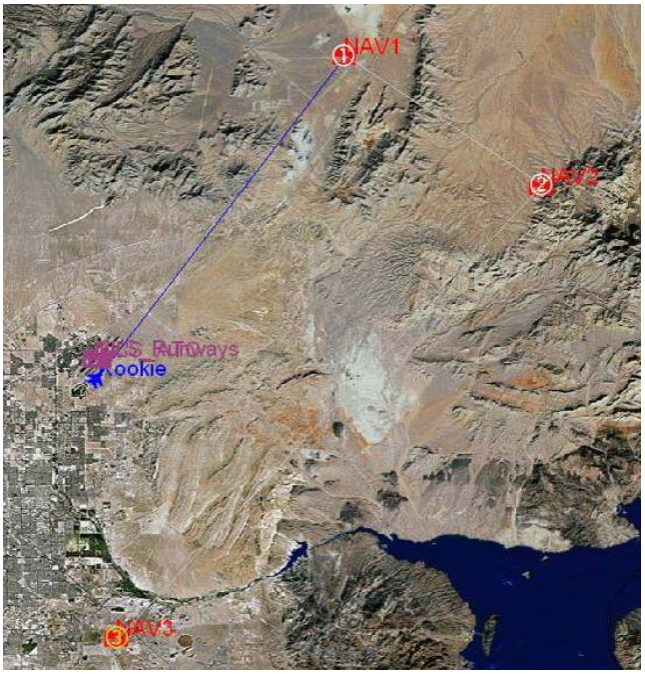


Figure 1. The navigation route in the AirBook® environment. The objective of the navigation task was to take off from an airport, navigate to three waypoints, and land at the initial airport.



(a) (b)

Figure 2. The Monitoring task required pilots to respond to the status of the master switch. In (a) the master switch is armed in (b) it is off.

Time Estimation. During the flight participants were required to estimate the duration of the task they performed (Zakay & Block, 2004). This prospective duration estimation can serve as a secondary task for workload measurement (Block, Zakay, & Hancock, 1999).

Subjective measures. Questionnaires included a demographic questionnaire, the DSSQ-S (Matthews, et al., 1999), and a computer version of the NASA-TLX (Hart & Staveland, 1988).

Procedure. After a brief description of the experiment by the experimenter, participants signed an informed consent form and completed a demographic questionnaire. Once informed consent was received, participants were familiarized with the AirBook® navigation task. They were given written and verbal instructions on how to maneuver through the AirBook® environment using the keyboard and the yoke provided for flight inputs. Then, participants had a five-minute practice. After the practice session, participants completed an initial NASA-TLX to be used as a baseline, and the DSSQ-S pre. Participants were then given the option to take a five-minute break. After this, participants were given brief instructions, and the dual task began. Participants navigated through a predefined route while performing the monitoring task for ten minutes. Performance on the monitoring task was recorded. In addition, during the dual task, participants received white noise through headphones. Half of the participants received noise at 55 dB, and the other half received noise at 85 dB. After completion of the dual task, participants estimated how long they were doing the dual task, completed a second NASA-TLX and the DSSQ-S post and were completely debriefed, having all their questions answered.

RESULTS

All analyses were conducted using SPSS 11.5 for Windows. All alpha levels were set to .05.

Flight Performance.

Navigation Task. Twenty-three of the 32 participants (72%) successfully completed the navigation task. Seven of those who did not complete the task had lower total flight hours than the mean. The other two had 140 and 700 flight hours, respectively. A one-way ANOVA on task completion with noise as a between subject variable was not significant, $F(1,30) = 0.146$, $p = 0.705$.

Monitoring Task. The mean reaction time for alarm 1 was 100.28 seconds (SD=186.36), and the mean reaction time for alarm 2 was 240.96 (SD=93.07). Twenty-eight of the 32 participants responded to alarm 1, and 23 of the 32 participants responded to alarm 2. All participants that failed to respond to alarm 1 had less flight hours than the mean except one participant, who had 700 hours. All participants that failed to response to alarm 2 had less flight hours than the mean except two participants, who had 140 hours and 700 hours.

A paired-samples T-test was conducted on the reaction times for alarm 1 and alarm 2 revealed that there was a significant difference in response time, $t(22) = -11.6$, $p < .0005$, Cohen's $d = -3.3$, effect size $r = -.86$. This difference is probably due to the fact that as the flight progressed participants paid less attention to the monitoring task. Two one-way ANOVAs on response to the first and the second alarm with noise as a between subject variable were not significant, $F(1,30) = 1.111$, $p = 0.3$, $F(1,30) = 0.146$, $p = 0.705$, respectively. The mean time to turn off alarm 1 in the 55dBA condition was 71.69 seconds (SD=138.9), and 128.88 seconds (SD=225.3) in the 85dBA condition. The mean time to turn off alarm 2 in the 55dBA condition was 208.75 (SD=94.5) seconds, and 276.09 (SD=81.4) seconds in the 85dBA condition. Two one-way ANOVAs on response time to alarm 1 and alarm 2 with noise as the between-subject variable were also not significant, $F(1,30) = 0.747$, $p = 0.394$, $F(1,21) = 3.322$, $p = 0.083$, Cohen's $d =$, effect size $r =$, respectively.

Stress and Workload.

NASA-TLX. A repeated-measures ANOVA before the task (M=52.40, SD=15.68) and after the task (M=53.25, SD=20.58) with noise as a between subjects variable revealed that the overall pre- and post- TLX measures were not affected by noise, $F(1,30) = 0.113$, $p = .739$.

A one-way ANOVA of the TLX scores after completion of the task indicated that there were no significant difference in NASA-TLX scores between the 55dBA condition and the 85dBA condition, $F(1,30) = 0.908$, $p = .348$. Additionally, the six sub-scales were analyzed for significance and effect size using one-way ANOVAs. The statistical results can be seen in Table 1.

Table 1. Analysis of the differences in TLX scores after completion of the dual task by noise condition .

NASA-TLX after completion of task	55dBA	85dBA	*F*	*df*	*p*	Cohen's *d*	Effect Size *r*
Global TLX	M=55.7 SD=21.5	M=50.8 SD=20.0	0.908	(30)	.348	.24	.12
Mental Demand	M=66.6 SD=24.4	M=58.4 SD=24.3	1.256	(1,30)	.271	.33	.16
Physical Demand	M=30.0 SD=18.3	M=35.5 SD=24.3	1.533	(1,30)	.225	-.25	-.12
Temporal Demand	M=56.9 SD=21.1	M=41.6 SD=25.5	0.058	(1,30)	.811	.65	.31
Performance	M=45.0 SD=33.2	M=36.3 SD=18.5	2.837	(1,30)	.103	.31	.15
Effort	M=70.3 SD=22.7	M=60.3 SD=30.0	2.034	(1,30)	.164	.38	.18
Frustration	M=35.6 SD=26.6	M=39.4 SD=30.9	0.244	(1,30)	.625	-.13	-.06

Time Estimation. The actual duration of the task was 10 minutes. The estimated times had a mean of 7.7 minutes (SD=3.0). Zakay (1998) suggested that underestimation is more likely to occur when time estimation is not the primary task at hand. As more attention is being focused on temporal information processing, more time signals are processed and the judgment is more likely to be accurate. Allocating fewer attentional resources to temporal information processing

causes fewer time signals to accumulate, resulting in a decrease in the estimated duration (Zakay & Block, 2004).

Time estimates were translated into the duration judgment ratio (*DJR; Block, Zakay, & Hancock, 1999*). Correlations were calculated for the DJR (M=77.3, SD=29.7) and the NASA-TLX score for participants in the 55dBA condition after the dual task, and the DJR and the NASA-TLX score for participants in the 85dBA condition to determine whether or not time estimates were correlated with noise condition. The correlation between the DRJ and the 55dBA condition was not significant, r = .164, p = .543. However, the correlation between the DJR and the 85dBA condition was significant, r = -.57, p<.05. This suggests that the 85dBA noise condition affected the DJR.

DSSQ-S. The DSSQ-S is scored into engagement, pre (M=21.3, SD=3.7) and post (M=22.7, SD=4.0), distress, pre (M=8.4, SD=5.8) and post (M=7.0, SD=4.4), and worry, pre (M=6.6, SD=4.6) and post (M=5.8, SD=4.3). Paired-samples t-tests were performed for engagement pre-post, distress pre-post, and worry pre-post, $t(31) = -1.97$, p = .058, Cohen's d = -.33, effect size r = -.17, $t(31) = 1.48$, p = .15, Cohen's d = .27, effect size r = .13, and $t(31) = 1.45$, p = .157, Cohen's d = .18, effect size r = .09, respectively. The difference in the pre and post engagement scores was not statistically significant, though it was very close. However, the effect size was not large enough to warrant more examination.

Repeated-measures ANOVAs were performed for each sub-scale with noise as the between subjects variable, $F(1,30) = 3.79$, p = .061 Cohen's d = .37, effect size r = .18, $F(1,30) = 2.15$, p = .153, Cohen's d = -.23, effect size r = -.11, and $F(1,30) = 2.104$, p = .157, Cohen's d = .06, effect size r = .03, respectively. Thus, noise was not a moderating condition for the DSSQ-S sub-scales.

Correlations were performed for the DSSQ-S measures before and after the task with noise condition and pre- and post- TLX global and performance scores using the Spearman ranking coefficient. For the correlations computed, the only ones that were statistically significant at the .05 level was the correlation between the post-task NASA-TLX global scores and the DSSQ-S post-task distress measure, ρ = .384, p = .03, pre-task distress and pre-task mental demand, ρ = .439, p = .012, pre-task distress and pre-

task frustration, $\rho = .434$, $p = .013$, and post-task distress and post-task effort, $\rho = .377$, $p = .033$. These correlations indicate that there is a strong relationship between overall workload, post-task effort and pre-task performance with distress after the task. Additionally, these correlations indicate that there is a strong relationship between pre-task mental demand and pre-task frustration with pre-task distress.

DISCUSSION

The majority of the pilots (71.9%) were able to complete the navigation task without any difficulties. However, since the workload estimates indicated that participants were not experiencing a heavy workload, one would think that close to 100% of the participants would have finished the task. Of the participants who could not complete the task, 4 were in the 55dBA condition and 5 were in the 85dBA condition. This suggests that task completion was not strongly affected by noise condition. Thus, task completion must be mediated by other moderators. When looking at the DSSQ-S engagement scores after the task, the mean was 22.6 compared to a maximum possible score of 28. Thus, the average score was only approximately 80% of the possible score. This suggests that participants were simply not engaged enough in the task to complete the task. To remedy this in the future, the navigation task will be comprised of more tasks that are shorter and a more demanding navigation scenario.

Performance on the monitoring task varied among participants, but this variation did not seem to correlate with the noise condition. The data suggests that some people simply ignored the alarms, possibly because there were no consequences, and the changes of the switch were very subtle. To better this task in the future, participants will be given a certain amount of time to comply with the alarm, and after that time expires, a visual alert will be shown to direct attention to the switch.

Interestingly, the mean time to turn off Alarm 2 in both conditions was more than twice the time it took to turn off Alarm 1. This suggests that participants were not monitoring the switch as closely throughout the flight as they were in the initial phases of flight. This is consistent with other monitoring task experiments, in which participants experience a performance decrement after a certain period of time.

The actual results were somewhat different from the expected results, in that workload, monitoring abilities, and subjective stress questionnaires were not affected by noise. However, as expected, task completion and time estimates were affected by noise. In the Hancock and Warm (1989) model, this means the stress was not sufficient to push participants out of the comfort zone, which is reflected in the lack of performance decrement in the flight parameters. Future work will address the issues that were problematic in this study.

ACKNOWLEDMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, Dr. Elmar Schmeisser, Technical Monitor, P.A. Hancock, Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect official Army or Department of Defense policy.

REFERENCES

Block, R.A., Zakay, D. & Hancock, P.A. (1999). Developmental changes in human duration judgments: a meta-analytic review. *Developmental Review, 19(1),* 183-211.

Chignell, M.H., & Hancock, P.A. (1987). A comparison of mental workload and available capacity in complex person-machine systems: A fuzzy-set theoretic approach. In: W. Karwowski and A. Mital (Eds.). *Applications of Fuzzy-Set Theory in Human Factors*. (pp. 271-288), Amsterdam: Elsevier Publishers.

Gluckman, J. P., Morrison, J. G., & Deaton, J. E. (1991). Complex task performance as a basis for developing cognitive engineering guidelines in adaptive automation. Proceedings of the Human Factors Society 35th Annual Meeting, Santa Monica, CA: Human Factors Society

Hancock, P.A. & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*, 519-537.

Hancock, P. A., & Weaver, J. L. (2005, *in press*). Temporal distortions under extreme stress. *Theoretical Issues in Ergonomic Science*.

Hildebrandt, M. & Harrison, M. (2002). The temporal dimension of dynamic function allocation. Paper presented at the *Eleventh European Conference on Cognitive Ergonomics.*

Parasuraman, R. & Mouloua, M. (1996). (Ed.). *Automation and human performance*. Lawrence Erlbaum: Mahwah, N.J.

Scallen, S.F. & Hancock, P.A. (2001). Implementing adaptive function allocation. *The International Journal of Aviation Psychology, 11(2),* 197-221.



Multi-Disciplinary University Research Initiative

MURI-OPUS

Operator Performance Under Stress

Pilots' Performance Under Stress

R. Nayeem, T. Oron-Gilad, & P.A. Hancock

MURI-OPUS Research Laboratory

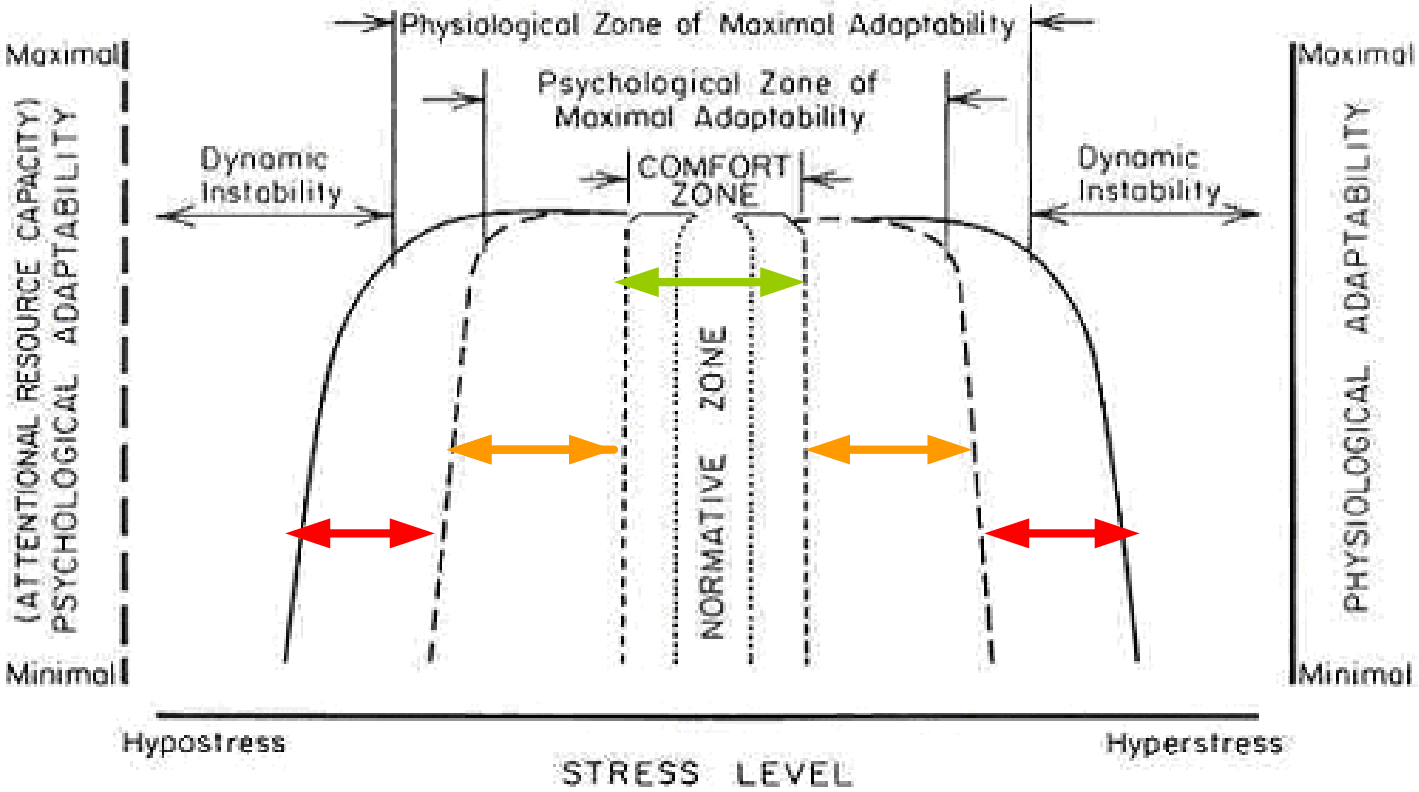
University of Central Florida

- Introduction
  - Temporal domain is stress-sensitive (Hancock & Weaver, 2005)
  - Narrowing of attention occurs - fixation
    - John Denver – Fuel selector handle
  - Time perception can be distorted (Block, Zakay,&Hancock, 1999)
  - Severe consequences can arise
    - Everglades – Eastern Airlines Flight 401

Hancock, P.A., & Weaver, J.L. (2005). Temporal distortions under extreme stress. *Theoretical Issues in Ergonomic Science*, **6** (2), 193-211.
Block,R.A., Zakay,D., & Hancock,P.A. (1999). Developmental changes in human duration judgments:A meta-analytic review. *Developmental Review*, **19**, 183-211.

- Introduction (cont'd)
  - Adaptive automation may help
    - Pilot's Associate
  - Function allocation is done automatically based on system state or human characteristics
  - Re-allocation should be done carefully (Hancock & Scallen, 1996)
    - Allocation Induced Oscillations

Hancock, P.A., & Scallen, S.F. (1996). The future of function allocation. Ergonomics in Design, 4 (4), 24-29.



Extended-U Model of Stress and Performance

Hancock, P.A., & Warm, J.S., (1989). A dynamic model of stress and sustained attention. *Human Factors*, **31**, 519-537.

- Purpose of the Study
  - Used white noise as stress to investigate performance and time perception on a dual navigation/monitoring task
- Expectations

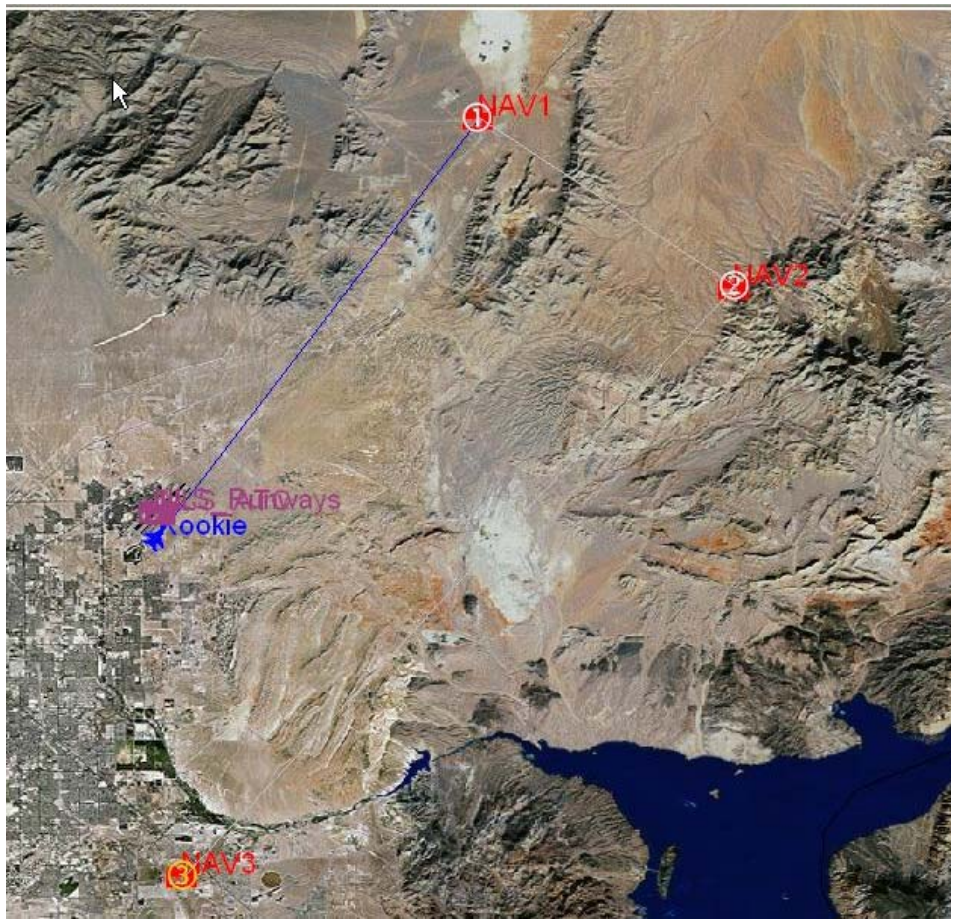
Measure	**Expectation when Stressed**
Workload	Increase
Task Completion	No change
Monitoring Abilities	Decrease
Prospective Time Estimation	Underestimate
DSSQ-S engagement	Increase
DSSQ-S distress	Increase
DSSQ-S worry	Increase

- Method
  - Participants
    - 32 total (28 males, 4 females)
    - At least private pilot's license
    - Total hours range from 40-700 (M=133.1, SD=134.1)
    - Undergraduate students from Embry-Riddle Aeronautical University
    - Compensated with extra credit

- Method (cont'd)
  - Apparatus
    - Computer
      - DELL laptop, 17" screen
    - Simulator
      - AirBook® by Simigon
    - White Noise
      - 55dBA & 85dBA through headphones
    - Flight Controls
      - GF flight stick



- Method (cont'd)
  - Navigation Task
    - Take off from airport
    - Navigate to 3 waypoints
    - Land at initial airport
    - 10 minutes



- Method (cont'd)
  - Monitoring Task
    - Monitor Master Switch
    - 30 seconds after takeoff and 30 seconds after flying over waypoint 2 switches to "off" position – participants must turn back on
    - No loss of power – simply monitoring task





- Method (cont'd)
  - Procedure
    - Informed Consent/Demographic Questionnaire
    - Practice navigation for 5 minutes
      - NASA-TLX & DSSQ-S Pre; 5 minute break
    - Dual task with noise for 10 minutes
      - NASA-TLX & DSSQ-S Post
      - Time Estimation
    - Debriefing

Matthews, G., Joyner L. A., Gilliland, K., Huggins, J. & Falconer, S. *Validation of a Comprehensive Stress State Questionnaire: Towards a 'Big Three'?* In I. Mervielde, I. J. Deary, F. De Fruyt & F. Ostendort (Eds.), (1999), Personality Psychology in Europe (Vol. 8). Tilsburg: Tilsburg University Press.

- Results
  - Navigation Task
    - 72% (23/32) of the participants completed the task
    - Seven of the nine that did not complete had fewer hours than the mean
    - One-way ANOVA on task completion with noise as a between-subjects variable (p = .705)
      - Noise did not affect task completion

- Results (cont'd)
  - Monitoring Task
    - 28/32 (88%) responded to Alarm 1 – all had total hours lower than the mean except 1
    - 23/32 (72%) responded to Alarm 2 – all had total hours lower than the mean except 2
    - Two one-way ANOVAs for response rate to alarms 1&2 with noise as the between-subjects variable (p = .3, p = .705, respectively)
      - Noise did not affect response rate

- Results (cont'd)
  - Monitoring Task (cont'd)
    - Alarm 1 RT (M=100.28s, SD=186.36s)
    - Alarm 2 RT (M=240.96s, SD=93.07s)
    - t-test on RT for Alarm 1 and Alarm 2
      - *$t(22) = -11.6$, $p < .0005$, Cohen's $d = -3.3$, effect size $r = -.86$*
      - There is a vigilance decrement
    - Two one-way ANOVAs on RT for Alarms 1&2 with noise as the between-subjects variable ($p = .394$, $p = .083$, Cohen's $d = -.76$, effect size $r = -.36$, respectively)
      - Noise did not statistically affect RT for Alarms 1&2

- Results (cont'd)
  - NASA-TLX
    - Workload did not change from pre-task to post-task
    - Repeated-measures ANOVA for pre- and post- global scores with noise as the between subjects variable (p = .739)
    - One-way ANOVA of the post-task TLX scores with noise as the between subjects variable (p = .348)
    - Sub-scale post-task scores were analyzed using one-way ANOVAs with noise as the between subjects variable
      - None were significant
      - Mental demand (.33), temporal demand (.65), performance (.31) and effort (.38) had moderate effect sizes
      - Noise did not statistically affect NASA-TLX scores

- Results (cont'd)
  - Time Estimates
    - Correlations between time estimate measures and 55dBA and time estimate measures and 85dBA (r = .164, p = .543, r = -.57, p<.05*, respectively)
      - At 85dBA, time distortion (underestimation) increased

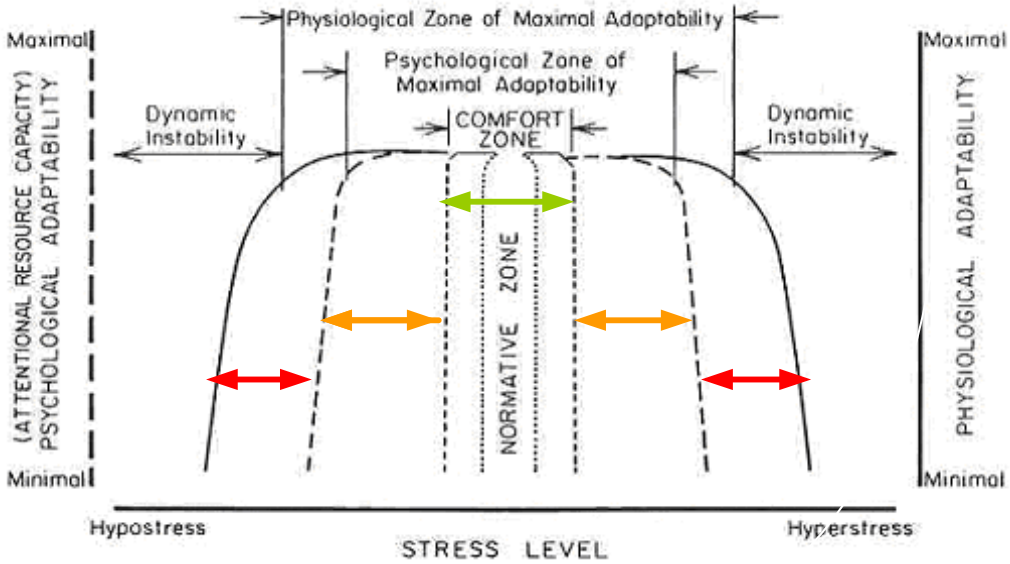
- Results (cont'd)
  - DSSQ-S
    - DSSQ-S scores were not different from pre-task to post-task
    - Correlations between DSSQ-S and noise condition and DSSQ-S and pre-task TLX scores and post-task TLX scores

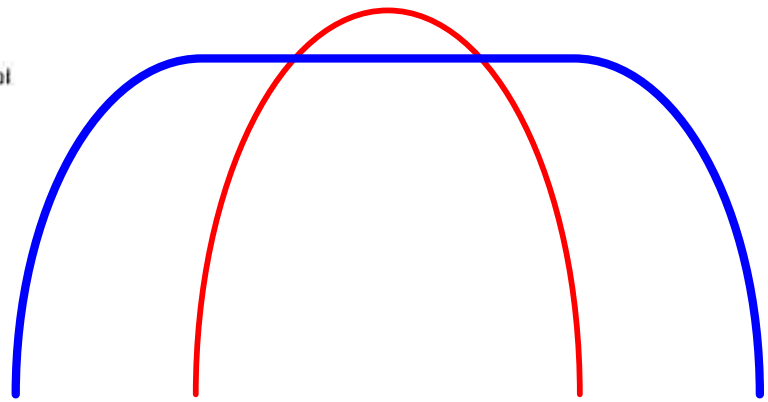
Measures	**Spearman's Rho *(ρ)***	**Significance *(p)***
Pre-task distress & Pre-task mental demand	.439	.012
Pre-task distress & Pre-task frustration	.434	.013
Post-task distress & Post-task global TLX	.384	.03
Post-task distress & Post-task effort	.377	.033

Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., Warm, J.S. (2002). Emotion. 2(4) 315-340

- Discussion

Measure	Expectation When Stressed	Actual
Workload	Increase	No change
Task Completion	No change	No change
Monitoring Abilities	Decrease	No change
Prospective Time Estimation	Underestimate	Underestimate – 85dBA underestimated more
DSSQ-S engagement	Increase	No change
DSSQ-S distress	Increase	No change
DSSQ-S worry	Increase	No change





Extended-U Model vs. Inverted-U Model

Hancock & Warm, 1989; Hancock, P.A., & Ganey, H.C.N. (2003). From the inverted-U to the extended-U: The evolution of a law of psychology. Human Performance in Extreme Environments, 7 (1), 5-14.

- Future Work/Improvements
  - Navigation task will have more tasks that are shorter and a more demanding scenario
  - Monitoring task will give visual alert if participant does not respond after a given time period to direct attention to that task



- Acknowledgements
  - This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, Dr. Elmar Schmeisser, Technical Monitor, P.A. Hancock, Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect official Army or Department of Defense policy.

The driving simulator as a workload research tool.

Justin F. Morgan, Tal Oron-Gilad, & Peter A. Hancock

University of Central Florida Department of Psychology, PO BOX 161390, Orlando, FL 32816-1390

Abstract

This paper shows that the driving simulator may be used as a platform in which to study workload issues. The participants were 5 adults with valid driver's licenses. After a familiarization drive participants drove a route with five equidistant and symmetrical segments using a map and turn-by-turn directions. Upon completion participants retraced their journey without the use of the map or directions. Measures of workload and simulation sickness were collected preimmersion and after completion of each of the two driving phases. No significant difference was found between the baseline familization drive and the driving with a map conditon. A significant difference was present between the driving with no map condition and both the baseline and map conditions. No significant results were present for the measures of simulation sickness. The findings support the use of driving and driving simulators as a workload research platform.

The driving simulator as a workload research tool.

Justin F. Morgan[1], Tal Oron-Gilad[1,2], and Peter A. Hancock[1,2]

1. The University of Central Florida,
2. Institute for Simulation and Training

Introduction

The task of driving is inherently stressful. There are large demands made of the driver's cognitive abilities, as well as physiological demands that controlling the vehicle require. Driving is generally considered a high workload task, as evidenced by the large amount of research performed examining the mental workload aspects of driving (for a review see Recarte & Nunes, 2003).

Increasing mental workload on the part of the driver is associated with longer reaction times to signals such as changing traffic lights and oncoming traffic (Patten, Kircher, Östlund, & Nilsson, 2004). In fact, many types of in-vehicle displays have been shown to lower overall driver performance (Lansdown, Brook-Carter, & Kersloot, 2004). In general, increasing drivers' mental workload has lead to a reduction in objective measures of performance.

Although there is often disagreement about the definition and boundaries of the topic, workload is one of the most commonly studied topics in human performance. Over the decades, many assessment techniques have been developed for measuring human workload. Common objective and subjective measures of workload take into account factors such as mental workload, temporal demands, and effort in a particular task to generate an overall description of workload. These measures have proven to represent the workload experienced by the operator in a variety of settings and conditions.

Although workload research has occurred in a variety of settings, little research has been performed which uses the concept of driving, and in particular driving simulation, as a platform for workload research. The present study seeks to provide data which supports the use of a driving simulator for workload research by demonstrating that reliable and accurate measures from a well-regarded workload assessment instrument are available in a simulated driving environment.

Method

Participants

Five adults (four males and one female) participated in this experiment. The five participants had an average age of 26.4 years. No particular inclusion/exclusion criteria were set for participation besides holding a valid driver's license, having normal color vision, and normal or corrected-to-normal vision. Participants were compensated for their participation with a payment voucher which can be exchanged for $7.50.

Simulator

One PatrolSim Mark III driving simulator was used for this experiment. The PatrolSim is a fixed-base, medium fidelity, driving simulator manufactured by L3 Communications (L3, 2005). It has all standard cabin controls of an automobile (steering, brake, throttle, turn signals, etc.), and sits in front of a projection screen which yields an area of approximately 150° horizontally which the participants were seated approximately three feet from.

Workload Measure

The NASA Task Load Index, or TLX, (Hart & Staveland, 1988) was administered via computer to each participant. The TLX is a multidimensional scale which measures six aspects of perceived workload (mental demand, physical demand, temporal demand, performance, effort, and frustration) on a scale from 0 to 100. In addition to the six individual scores and their weights, an overall value of perceived workload is generated. For the purpose of this research, only the combined, total workload score is examined.

Simulation Sickness Measure

Developed by Kennedy, et al. (1993), the Simulator Sickness Questionnaire (SSQ) is a widely used multidimensional measure of simulation sickness. The SSQ uses a pre-immersion interview and symptom checklist as a baseline to which subsequent symptom checklists are compared to. The SSQ measures symptoms clustered into three dimensions: nausea, oculomotor, and disorientation.

Procedure

After obtaining informed consent, the participant was administered the SSQ pre-immersion interview and symptom checklist forms. Following completion of these, the participant completed a baseline measure for the NASA TLX. Participants were then asked to enter the simulator. Each participant was first given a five minute period to familiarize them with the operation of the simulator. This initial exposure to the simulator was done in order to reduce the likelihood of the participant experiencing symptoms of simulator sickness and to reduce the participant's uncertainty with the control of the simulator.

Participants were then given a map and a set of turn-by-turn driving directions. Participants were instructed to proceed according to the directions to a series of six markers placed at approximately equal distances in the environment. When they reached each marker, participants were instructed to come to a complete stop, honk the automobile's horn, and then continue to the next marker. Participants were informed that they could refer to the map and driving directions at any point in the task. After completing the driving route, post task versions of the SSQ and NASA TLX were administered.

For the second driving task, participants began from the last marker and were asked to navigate through the markers in reverse order, without the aid of a map or driving directions. Participants were instructed to continue as far as possible and that

when they felt they were lost or unable to find the next marker, inform the experimenter of this and the experiment would stop. At completion of the route or after the participant could not find the next point, post task versions of the SSQ and NASA TLX were administered and the participant was debriefed.

Results

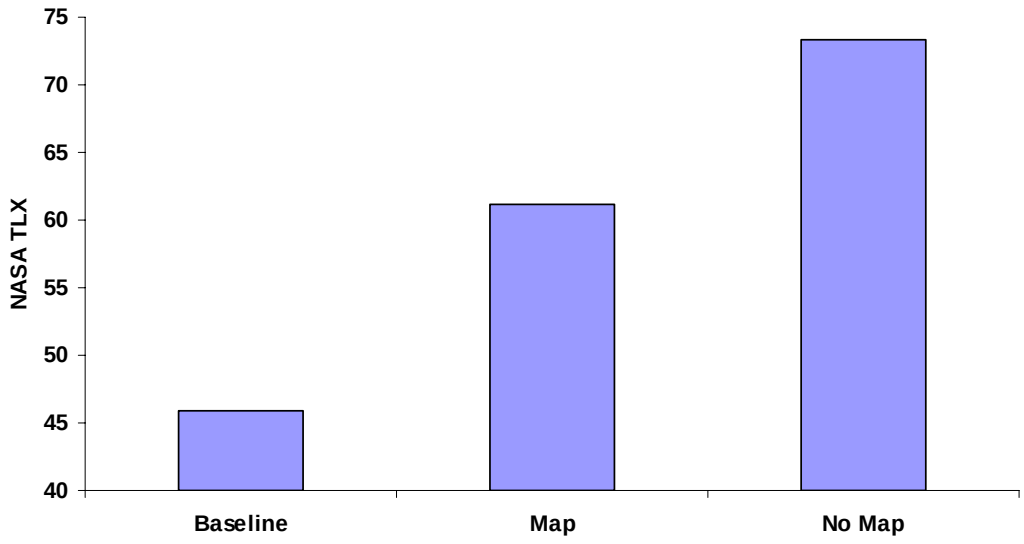
Time Between Markers and Time To Complete Phases

A repeated measures ANOVA demonstrated no significant differences between the individual segments of the journey between markers, $F(4,4) = 1.67$, $p = .32$. Data from all five segments were collapsed for comparison of the two conditions. A paired samples two-tailed t-test showed that there were no significant differences between the map and no map conditions, $t(1) = 2.86$, $p = .214$. Means for the two phases were $M = 2.10$, $SE = .31$ for the map condition, and $M = 1.10$, $SE = .04$ for the condition without a map.

NASA TLX

A significant change was present in the workload scores across the three phases (practice session, driving with map, driving without map), $F(2,12) = 7.15$, $p = .01$. Planned comparison analysis shows that the map ($M = 61.20$, $SD = 6.46$) and no map ($M = 73.33$, $SD = 5.20$) phases differed significantly from one another, while the baseline ($M = 45.93$, $SD = 18.08$) differed significantly only from the no map phase.

Figure 1



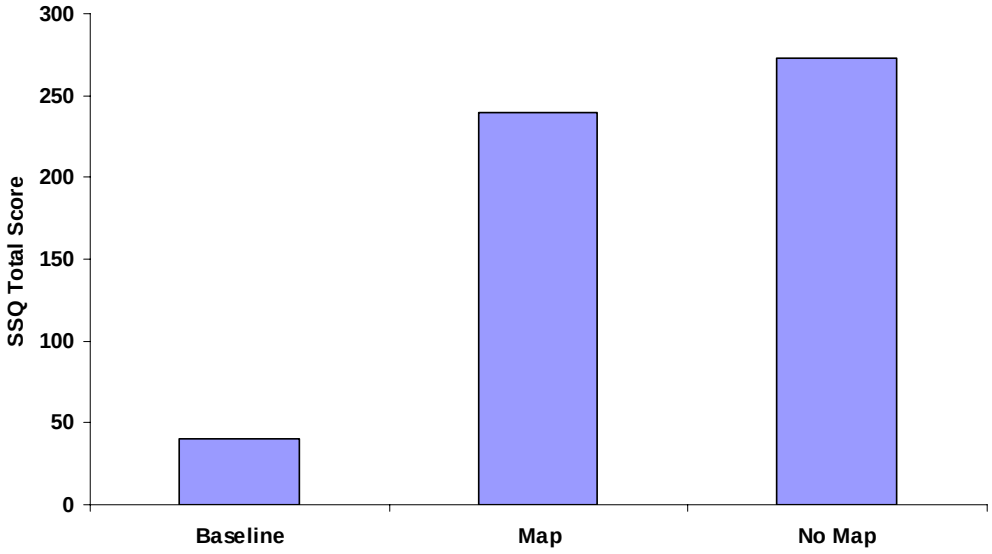
NASA TLX Scores by phase of participant drive

SSQ

Although there were no significant effects were present for the data from the SSQ, $F(2,12) = 3.399$, $p = .068$, an increase in SSQ total scores was observed between the

baseline measure (*M* = 40.77, *SD* = 91.16), the first post-immersion measure after the map condition (*M* = 239.55, *SD* = 128.17) and the last measurement after the no map condition (*M* = 273.18, *SD* = 211.86).

Figure 2



SSQ Scores by phase of participant drive

Discussion

The lack of any significant differences between time measures for driving different segments of the route demonstrated that the route was symmetrical and equidistant in layout. This allowed for the examination of the workload measures as a measure of the navigational task instead of a measure of the driving task. The lack of a significant difference between baseline and map driving conditions on the workload measure show that the actual task of driving, when no additional stressors are introduced, is not a significant contributor to workload. The addition of an additional workload item, in this case the navigation without a map, greatly increases workload over both baseline and map conditions. These support the hypothesis that the navigational task was the significant contributor to workload in this experiment.

Participants, on average, did not experience many symptoms of simulator sickness during the experiment. This is supported by the gradual, yet non-statistically significant, increase in scores on the SSQ. The lack of a significant increase in the SSQ leads to the conclusion that workload was not being influenced by any symptoms of simulator sickness, and was instead influenced solely by the navigational task.

Of interest in this research project is the validation of a ready to use driving simulator for workload research. We attempted to demonstrate this by showing that while driving an equidistant and symmetrical route under varying degrees of workload, high quality measures of workload, which were due to the task characteristics instead of the act of driving, could be obtained. This was supported by the experimental findings, and

supports the idea that a driving simulator can be used as a platform for workload research...

References

Hart, S.G. & Staveland, L.E. (1988). Development of the NASA TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. International Journal of Aviation Psychology, 3, 203-220.

L3. (2005). STS Police/Law enforcement driving simulators. Retrieved July 5, 2005 from http://www.shipanalytics.com/STS/patrolsimii+.asp

Lansdown, T.C., Brook-Carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. Ergonomics, 47, 91-104.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S. E., Falconer, S. & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state 'Big Three'? In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7,pp. 335-350). Tilburg, the Netherlands: Tilburg University Press.

Patten, C.J.D., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: Cognitive workload and attention resource allocation. Accident Analysis & Prevention, 36, 341-350.

Recarte, M.A., & Nunes, L.M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. Journal of Experimental Psychology: Applied, 9, 119-137.

Road Environment and Passive Fatigue

Tal ORON-GILAD, & Peter A. HANCOCK
Department of Psychology and Institute for Simulation and Training
University of Central Florida, Orlando, FL

Summary: We distinguish between fatigue caused by the demands of the driving task itself (see Hancock & Desmond, 2001) from the standard traditional approach that links fatigue predominately to the lack of sleep. Fatigue can be caused by two sources; 1) the driver's initial state before starting the drive, 2) the characteristics of the drive and the road environment, and both sources can have a cumulative effect. It is not clear what principles which are involved in making one road environments more prone to inducing driver fatigue than another. For the purpose of the current presentation we provide empirical data on road environment and driver fatigue summarized from a series of three experiments that the first author has conducted at Ben-Gurion University (see Oron-Gilad, 2003; Oron-Gilad, Ronen, Shinar, & Cassuto, 2001). Those are examined in relation to the Hancock and Warm (1989) model of adaptability. The most significant and consistent findings of the three experiment is in the way that fatigue is reflected in driving performance across different road environments. These findings suggest that drivers are flexible in the way they handle fatigue over the course of time. They can adopt different strategies to compensate for their performance decrement, by focusing efforts on critical elements of each different type of roadway. Understanding of this dependency of fatigue symptoms on road conditions is of especial relevance to designers of technological fatigue countermeasures as well as those of future roadway systems.

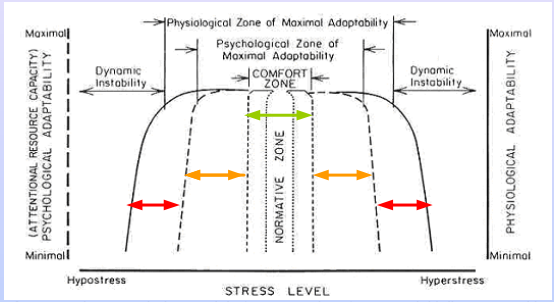


Figure 1. The extended-U model of stress and performance initially proposed by Hancock and Warm (1989).

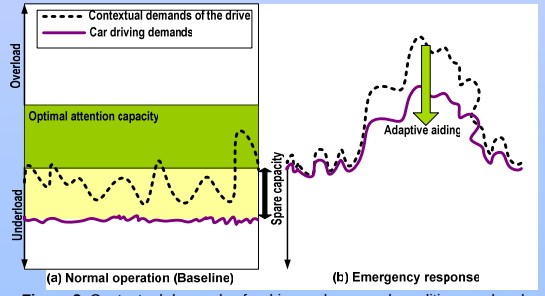


Figure 2. Contextual demands of a drive under normal conditions and under emergency responses. In case of an emergency, drivers may need adaptive aiding to compensate for lack of attentional capacity.

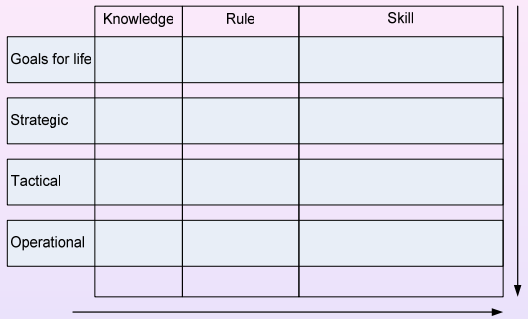


Figure 3. A matrix representation of driving components and their complexity of demand. Uncertainty decreases as the degrees of freedom are reduced. For example, a professional truck driver does not have the option to choose not to drive to a certain location at a certain time which may reduce the flexibility in the highest 'goals for life' level.

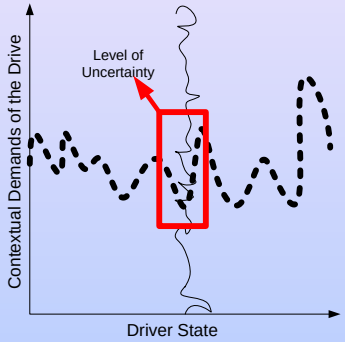


Figure 4. Uncertainty of the drive is a function of driver state and the contextual demands of the drive. Smaller fluctuations occur in driver state throughout a drive than in the contextual demands of the drive.

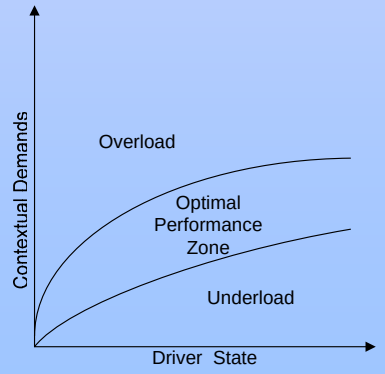


Figure 5. Based on the Hancock and Warm (1989) model, uncertainty can be grossly categorized into three categories: underload, optimal and overload.











Figure 6. (a) The driving simulator at BGU (b) A mandatory-service military truck driver participating in the road characteristics study (c) The driving scenario consisted of three road segments, each simulating a different road in the southern part of Israel. A winding road with 22 curves on a 9.2 Km long segment {1}, driven either uphill {1a} or downhill {1b}, a two-lane straight 13.3 Km rural road {2}, and a straight four-lane divided highway with low traffic density 13.0 km long {3 (d) Real life scenery and (e) Is he sleeping? He just started to drive…

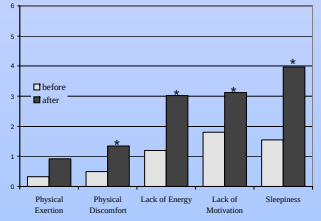


Figure 7. Subjective Occupational Fatigue Inventory (SOFI, AAhsberg 1998) before and after the prolonged drive (all drivers).
* Difference significant at p<.02

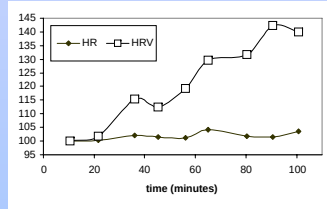


Figure 8. Average level of HR and HRV during the drive relative to the initial level (100).

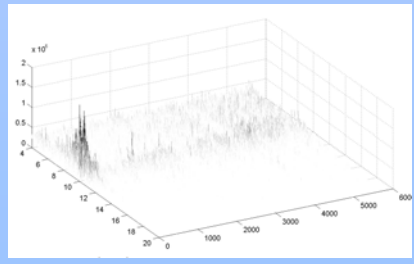


Figure 9. The decrease of alpha power at the beginning of the drive (as shown in detail in Picture 1) and an increase of power as the drive proceeds. There is also an increase of the power of theta waves as the drive proceeds. The power of the beta waves remained constant throughout the drive.

References
Hancock, P. A. and Warm, J. S.(1989). A dynamic model of stress and sustained attention. Human Factors, 31, 519-537.
Hancock, P.A., & Desmond, P.A. (Eds.). (2001). *Stress, workload and fatigue.* Lawrence Erlbaum, Mahwah: NJ.
Oron-Gilad, T. (2003). *Alertness maintaining tasks while driving.* Unpublished doctoral dissertation, Ben-Gurion University, Beer-Sheva, Israel.
Oron-Gilad, T. (Submitted) Countermeasures to Passive Fatigue and Underload in Driving - A Theoretical Perspective, *Human Factors*
Oron-Gilad, T., & Shinar, D. (2000a). Driver Fatigue among Military Truck Drivers. Transportation Research part F, *3*(4), 195-209.
Oron-Gilad T., Ronen A., Shinar D. & Cassuto Y. (2001). Fatigue of professional Truck Drivers in Simulated Driving: A preliminary Study, *Proceedings of the Traffic Safety on Three Continents International Conference,* Moscow, Russia.

Comparison of Spatial and Temporal Discrimination Performance across Various Difficulty Levels

J. E. Thropp, J. L. Szalma, and P. A. Hancock
University of Central Florida

ABSTRACT

Spatial and temporal information may be perceptually related, such that performance on spatial and temporal discrimination tasks may draw upon common resource capacities. However, the degree of this relationship is not clear. Multiple levels of difficulty in spatial and temporal discrimination tasks were manipulated in the current study in order to examine the perceptual mechanisms of both dimensions and how they relate to one another. It was found that performance across difficulty levels differed for the spatial and temporal tasks, such that although both tasks showed a linear relationship between performance and discrimination difficulty, performance was superior in the spatial condition. Furthermore, these dimensions cannot be easily equated even psychophysically, and they may not share common perceptual mechanisms as previously suggested.

INTRODUCTION

The Hancock and Warm (1989) model of human adaptation under various levels of stress and task load established the importance of understanding the spatial and temporal characteristics of tasks. Within the context of that model, the spatial elements are instantiated in the information structure dimension and the temporal elements are captured in the information rate dimension (see Figure 1). However, what is not clear is the degree to which these dimensions are related perceptually, and how changes along

one dimension affect an individual's position along the other dimension. It has been suggested (Hancock, Szalma, & Weaver, 2002) that information processing demands along these dimensions may draw on common resource capacities. However, recent studies presented at this conference have indicated that the relation between these dimensions may not be so simple (Ross, Szalma, Thropp, & Hancock, 2003; Thropp, Szalma, & Hancock, 2004). For instance, in a perceptual discrimination study deriving performance operating characteristics to examine resource sharing between spatial and temporal tasks, Thropp et al. (2004) reported that the functions obtained varied substantially across participants. In the current study we sought to examine the relations between these dimensions psychophysically by comparing multiple levels of discrimination difficulty across the two dimensions. It is hoped that this will provide the basis for examining the perceptual mechanisms underlying spatial and temporal processing.

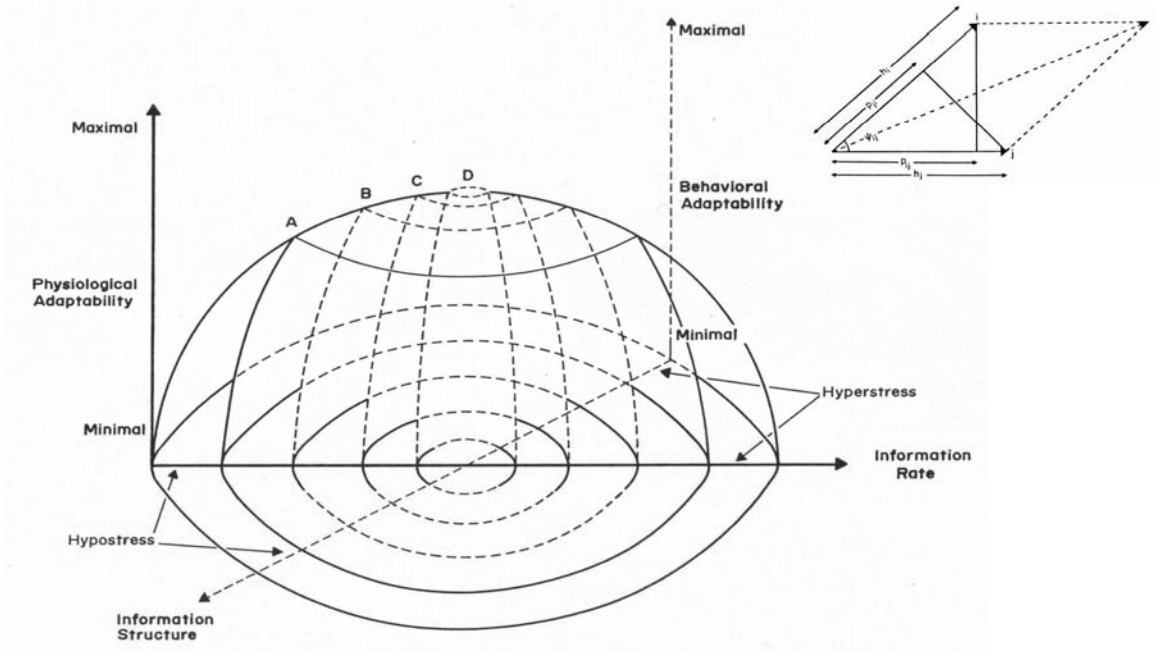


Figure 1. Adaptation under Stress. After Hancock & Warm (1989)

METHOD

Participants

Participants were 28 undergraduate students (12 males, 16 females) enrolled at the University of Central Florida, all with normal or corrected to normal vision. Participants were compensated by payment of $15 for an experimental session lasting one hour and forty five minutes.

Experimental Design

Observers engaged in both a spatial and a temporal visual discrimination task, and experienced seven levels of discrimination difficulty within each task type. This yielded a 2 (task type) by 7 (discrimination difficulty) within-subjects design. The spatial discrimination task required participants to identify which of two vertical lines was taller. A trial consisted of presenting each stimulus sequentially, followed by a response screen prompting a decision by the observer regarding whether the taller of the two lines was presented first or second. The 'shorter' line was held constant at 32mm across the levels of difficulty. The difficulty of the discrimination was manipulated by increasing the 32mm line by increments of .5, 1.0, 1.5, 2.0, 2.5, 3.0, or 3.5mm. In the temporal discrimination task, participants were required to determine which of two lines, presented sequentially, was longer in duration. The 'shorter' duration was held constant across conditions at 300ms, and the difficulty of the discrimination was manipulated by increasing the presentation time of the line by increments of 10ms, 40ms, 70ms, 100ms, 130ms, 160ms, and 190ms. For both tasks, participants responded by pressing the '1' key on the keyboard if the first line was the tall line (in the spatial discrimination task) or

the line was presented for the longer duration (in the temporal discrimination task), or by pressing the '2' key if the second line was the tall line or presented for the longer duration. For each of seven difficulty conditions observers experienced 30 trials. The 'taller' or 'longer' stimulus was presented first in 15 trials and second in the other 15 trials in each 30-trial block. Participants completed the spatial and temporal tasks in blocks. Half of the participants experienced the spatial block first and the other half experienced the temporal block first. The orders in which the difficulty conditions were presented varied across participants and were determined by using a Latin Square.

RESULTS AND DISCUSSION

The mean proportion correct scores for the spatial and temporal discrimination tasks are displayed as a function of difficulty level in Figure 2. It can be seen in the figure that the performance varies from near chance (.5) in the most difficult (i.e., smallest change to be discriminated) conditions to approximately .9 in the easiest conditions. In addition, participants achieved better performance in the spatial than in the temporal task. These impressions were confirmed via a 2 (task) by 7 (difficulty) ANOVA, which revealed a significant effect for task $F(1, 27)=14.687$, $p<.005$, with spatial task performance (M=.778) higher than temporal task performance (M=.711). There was also a significant effect for difficulty level across both tasks $F(6, 162)=88.662$, $p<.0001$, such that performance decreased with increasing difficulty levels. Mean proportion correct for the most difficult discrimination condition through least difficult condition were M=.548, .642, .708, .739, .814, .873, and .887, respectively. Post-hoc analyses revealed that proportion correct on all difficulty levels differed significantly from each other ($p<.05$), except that performance on the third most difficult task did not differ significantly from

the fourth most difficult task and performance on the sixth most difficult task did not differ significantly from the seventh most difficult task ($p>.05$ in each case). In addition, a trend analysis indicated a significant linear trend across difficulty levels $F(1, 27)= 357.489$, $p<.0001$. Thus, there is a linear relationship between performance and degree of difficulty.

The results of this study indicate that the spatial and temporal dimensions cannot be easily equated even psychophysically, and they may not share common perceptual mechanisms as suggested previously by Hancock and his colleagues (Hancock, Szalma, & Weaver, 2002; Hancock & Weaver, 2005). Further work is needed to evaluate whether changes in performance along one dimension influence performance on the other dimension. However, the current results suggest that the performance effects of increasing task load may be linear for both the information structure and information rate dimensions of the Hancock and Warm (1989) model. In addition, results of both this experiment and that of Thropp et al. (2004) revealed a substantial degree of individual differences in performance on spatial and temporal tasks. For instance, some individuals tend to perform better on one task than the other, while other are more consistently high or low in performance levels on both tasks. Subsequent analyses will evaluate the potential influence of several theoretically relevant individual difference variables in order to further explore the factors controlling relative performance along these dimensions.

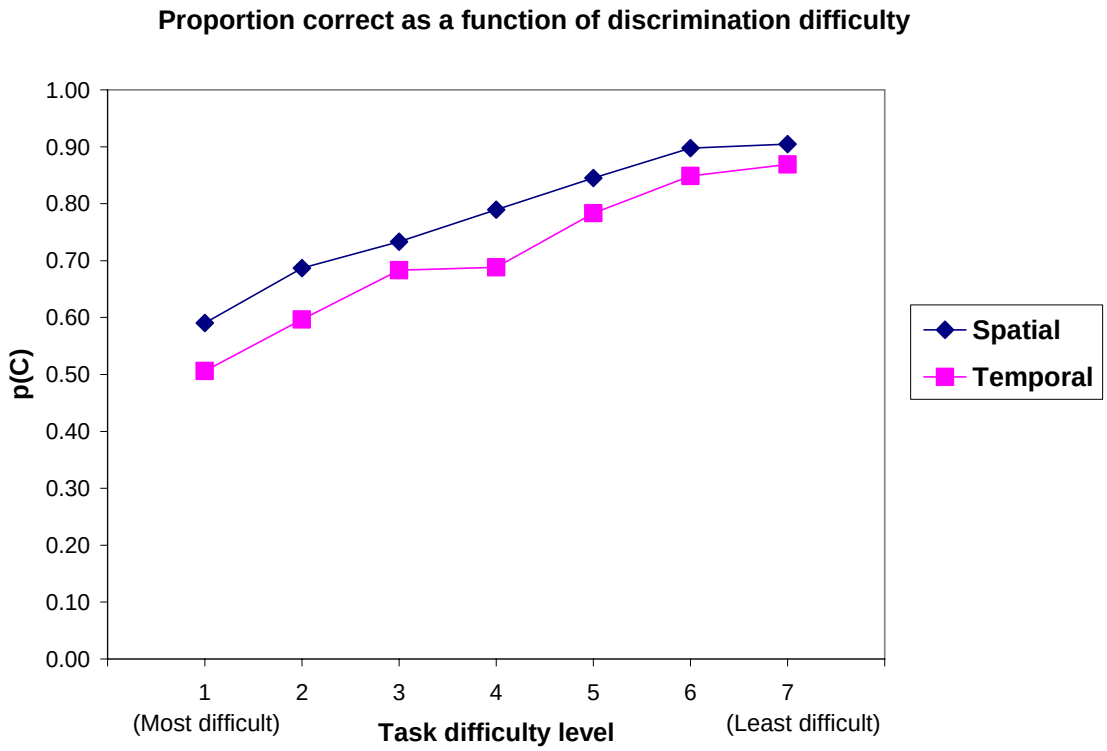


Figure 2. Mean proportion scores for each difficulty level in spatial and temporal tasks.

REFERENCES

Costa, P. T. & McCrae, R. R. (1992a). *NEO PI-R Professional Manual,* Odessa, FL: Psychological Assessment Manuals.

Eysenck, M.W. (1992). Anxiety: The cognitive perspective. Hillsdale, NJ: Erlbaum.

Eysenck, H.J. & Eysenck, M.W. (1985). Personality and individual differences: A natural science approach. New York: Plenum Press.

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4,* 26-42.

Hancock, P. A. & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519-537.

Hancock, P.A., Szalma, J. L. & Weaver, J. L. (2004). The distortion of perceptual space-time under stress. *Proceedings of the 23rd Annual Army Science Conference*, Orlando, FL (IP-23).

Hancock, P. A. & Weaver, J. L. (2005). On time distortion under stress. *Theoretical Issues in Ergonomics Science, (6)2,* 193-211.

Matthews, G., & Campbell, S. E. (1998). Task0induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 821-825.

Matthews, G., Coyle, K., and Craig, A. (1990). Multiple factors of cognitive failure and their relationships with stress vulnerability. *Journal of Psychopathology and Behavioral Assessment, 12*, 49-64.

Matthews, G., Deary, I.J., & Whiteman, M.C. (2003). Personality traits. Cambridge: Cambridge University Press.

Ross, J. M., Szalma, J. L., Thropp, J. E., & Hancock, P. A. (2003). Performance, workload, and stress correlates of temporal and spatial task demands. *Proceedings from the Annual Human Factors and Ergonomics Society Meeting, 47,* 1712-1716.

Thropp, J. E., Szalma, J. L . & Hancock, P. A. Performance Operating Characteristics for Spatial and Temporal Discriminations: Separate or Common Capacities? Presentation given September 24, 2004 at 48th Annual Meeting of the Human Factors and Ergonomics Society, New Orleans, LA.

WHY AND HOW HFE PROFESSIONALS CAN BETTER USE THEORY (METATHEORY INCLUDED; SOME ASSEMBLY REQUIRED)

Mark E. Koltko-Rivera[1] and Peter A. Hancock[2,3]

[1]Professional Services Group, Inc., Winter Park, FL

[2]MURI-OPUS Laboratory
[3]Institute for Simulation and Training
University of Central Florida
Orlando, FL

The relationship between theory and discipline is problematic for human factors and ergonomics (HFE). We address the following constituent issues: (a) the present state of theory usage in HFE; (b) the reasons underlying this state; (c) the need for theory in HFE; (d) what HFE professionals (including educators, journal editors, and individual researchers) should do to encourage the proper use of theory; and, (e) the outlines of a metatheory of HFE. A metatheory is a general framework that may help professionals to construct more useful specific theories. Our metatheory of HFE involves five basic units, or classes of variables: task, environment, personnel, tool, and performance. Each unit in turn has multiple components (i.e., specific variables). Use of the metatheory is illustrated with specific examples. Our hope is that researchers will be motivated to make explicit and useful connections between their research and necessary theory, to the improvement of both; the metatheory may be useful in this endeavor.

INTRODUCTION

Let us define a *theory* as a particular set of constructs and/or observables, and the relationships that may be proposed to exist among them. It has been repeatedly claimed by professionals in a variety of sciences, basic and applied, that theory is fundamental and crucial to the successful pursuit of their disciplines (e.g., Kerlinger & Lee, 2000). However, the relationship between theory and discipline is somewhat problematic in the case of human factors and ergonomics (HFE). In this work, we address the following issues:

- the present state of theory usage in HFE;
- reasons underlying the present state of theory usage in HFE;
- the need for theory in HFE;
- what HFE professionals should do to encourage the proper use of theory in research;
- the outlines of a metatheory of HFE.

PRESENT STATE OF THEORY USAGE IN HFE

Although there are exceptions (e.g., much of what is published in *Theoretical Issues in Ergonomics Science*), we believe that it is fair to say that the present state of HFE theory usage, as expressed in the research literature, is poor at best. We say this for two reasons.

First, much HFE research is based on very-small-scale implicit theories that are of limited applicability; this has unfortunate consequences for the generalizability of the research. Second, much, perhaps most, HFE research attempts to make no explicit contribution to theory, nor is a fair proportion of such research designed to do so.

REASONS UNDERLYING THE STATE OF THEORY USAGE IN HFE

There are probably several reasons for the current state of theory usage in HFE. We focus on five here.

First, some HFE professionals seem to have the sense that HFE is "about applied research—period." From this point of view, theory is peripheral or irrelevant to the HFE enterprise.

Second, there appears to be a lack of formal training in theory building within HFE programs in academia. Of course, this is a problem that is endemic to many disciplines and sub-disciplines, and not to HFE alone.

Third, there seems to be a widespread sense that HFE can rely on other disciplines to devise theory for the use of HFE professionals. Supposedly, these other disciplines are better equipped to deal with the rigors of theory-building.

Fourth, a vicious cycle appears to be at work. Usually, HFE professionals do not write theory; the body of theory then may lack obvious relevance to HFE; HFE professionals then are less inclined to use or write theory; and so on.

Fifth, HFE professionals, when they use theory at all, tend to rely on theories of very limited scope. This is the case, even when much larger-scale theories might be much more appropriate. Thus, we might focus on theories of cockpit design, even when more general theories of human cognition might be highly relevant and rewarding if used.

NEED FOR THEORY IN HFE

There are several reasons why HFE professionals should develop and test theory, and seek to make contributions to theory through their applied research. We focus here on two:

the existential situation, and the epistemic status, of the HFE professional.

HFE Professionals Are Already Deeply Engaged With Theory—They Just Have Not Recognized That Yet

For the most part, even the most applied HFE research involves not only observable characteristics (e.g., size, mass, acceleration, reaction time, errors, etc.), but also characteristics that are not directly observable, that is, *constructs*. (Thus, objectives, motivation, mental workload, and even the notion of "task," are all constructs). Once HFE professionals—or any scientific or technical professionals—find themselves in the position where they need to consider relationships between constructs, and between constructs and observables, these professionals have already entered deeply into the territory of theory. This is because the very notion of theory involves a set of constructs, their definitions, and the relationships between them (see above; Kerlinger & Lee, 2000, p. 11). If you work with constructs, you are already working with theory; as the saying goes, pick up one end of a stick, and you have just picked up the other end as well.

The issue, then, is not, must we use theory in our research? That bridge has already been crossed whenever we encounter constructs, in even the most applied research; engagement with theory is then inescapable, and has already occurred (see Hancock, Weaver, & Parasuraman, 2002). We might as well try to conduct construct-laden HFE research outside the boundaries of space and time, as well as we could try to conduct it without theory. The issue is rather, do we proceed with implicit and poorly conceived theories, or do we make our theories explicit, well-constructed, and representative of the real world?

HFE Professionals Are the Best Potential Builders of HFE-Relevant Theories

Other disciplines should not be expected to have the needs or knowledge of HFE professionals in mind when developing theory. HFE professionals are, in many ways, better situated than professionals in other disciplines to develop theory that is relevant to HFE concerns. The problem is that many HFE professionals (judging from the literature) either do not bother with theory, or do not use it effectively, perhaps as a result of the evident lacunae in theory (see our comments above concerning a vicious cycle).

Theory construction proceeds best in a context where one is thoroughly familiar with the facts of the phenomena under investigation. Who is more familiar with the facts of HFE-related phenomena than HFE professionals? Add to this phenomena-related knowledge base another knowledge base regarding theory development and use, and HFE professionals have the potential to make substantial contributions to the overall theoretical edifice of science.

ENCOURAGING THE USE OF THEORY IN HFE

There are several actions that professionals within HFE can take to promote the proper use of theory within HFE. These include actions by educators, journal editors, and individual HFE researchers.

What Educators Can Do

Educators can encourage a theory-driven perspective among future HFE professionals. There are several forms that this encouragement might take (the following not being mutually exclusive).

First, a course in theory-building can become an integral part of the curriculum. (We would argue for this being a required course.) The burden that this extra teaching load imposes on faculty can be lessened by recognizing that, in fact, all sectors in any discipline can benefit from training in theory building. Thus, not just the HFE program, but an entire Department of Psychology/Engineering/Business/Cognitive Science, etc., can offer a course in theory building, inviting all majors across the discipline. Responsibility for teaching this course can thus be shared or rotated across all faculty within a department. (For suggestions on texts, see below.)

Second, at the department level, policy can be set to ensure that many required courses include a focus on the building of theory and its use in research. Focusing on the issue across courses emphasizes its importance.

Third, theory building and critique can be a component of the doctoral candidacy process. This can be done regardless of whether a department has a portfolio- or an examination-oriented approach to establishing candidacy.

What Journal Editors Can Do

Journal editors can adopt the policy that authors of research articles must explicitly tie their work to theory, as some presently require a tie to an application area or a design. (It may be necessary for journals to have tutorial articles demonstrating better and worse ways in which to do this.)

Journal publication policies can have a strong effect on the literature. For example, the American Psychological Association (APA) strongly encourages all authors of quantitative research reports submitted to APA journals to justify their sample sizes on the basis of formal considerations of statistical power (American Psychological Association, 2001, p. 24). Although compliance with this stipulation is far from perfect, it would appear that the number of authors in APA journals who pay attention to statistical power requirements has risen substantially. The same can happen with the explicit tying of research to theory.

What Individual HFE Professionals Can Do

Most importantly, researchers can do several things to relate their work to theory in useful ways. These include principles of research or authorial practice, and self-education.

The first author of this paper, in teaching courses in research design, has made it a point to convey the notion that any research project relates to at least two questions: the Smaller Question and the Larger Question. The Smaller Question is the specific matter in front of the researcher (e.g., which of two cockpit displays results in fewer pilot errors?). The Larger Question relates to more fundamental issues (e.g., what sorts of task-relevant circumstances make for accurate versus error-prone performance?). We contend that any given subdiscipline (e.g., HFE, personality psychology, structural engineering) has about a dozen or so Larger Questions that

dominate the subdiscipline. (We suggest that it would be useful for HFE journals to run special issues delineating the Larger Questions in the field, rather as the content areas have been specified in *Human Factors*.) Individual researchers can make it a point to situate their research explicitly in relation to these two questions, from the earliest stages of planning their research right through the composition of their reports.

Researchers should state explicitly their theoretical frameworks several times: (a) to themselves, before they develop their research, so as to make their assumptions clear, an activity that may lead to redesign; (b) to themselves again, after they conduct their research and obtain findings, with the idea that their theory may need to be modified; (c) to the reader of their research report (perhaps in "before" and "after" versions, depending on their research results). Such practices will help to bring theory to the fore in a clear way.

In tying research to theory, it should be noted that the use of middle-to-large-scale theories is preferable to the use of tiny-to-small-scale theories. This is because the larger the scale of the theory, by definition, the more the theory connects to a larger body of knowledge within the discipline and across disciplines. In turn, the more we do this, the more generalizable our research becomes.

Researchers should seek to educate themselves regarding the use, development, and critique of theory. Many of our doctoral programs provide insufficient education on theory in general, instead focusing on specifically dissertation-related theory. We learn various theories, but possess insufficient information regarding how to construct new ones; all too often, we do not learn much about how to evaluate or modify theories, in the face of new data, to improve these theories' utility. Consequently, much effort must be expended in self-education. Fortunately, there are resources available for this task. Excellent places to begin include the works of Dubin (1978) and Kukla (2001). There is much to be learned from publications that focus on the need for a critical stance concerning the assumptions underlying any given theory (e.g., Slife, Reber, & Richardson, 2005; Slife & Williams, 1995).

OUTLINES OF A METATHEORY OF HFE

For the benefit of HFE professionals, we here outline a sketch of a metatheory of HFE. Although a full exposition of this theory must await a larger treatment in a journal article (pending; Koltko-Rivera & Hancock, 2005), we present this here with the hope that this might assist individual theorists to consider the larger frameworks in which they might position their work theoretically.

"Metatheory" Defined

A metatheory may be thought of as a very general framework that in turn can generate many specific theories. Although not itself a theory, a metatheory may help professionals to construct specific theories more carefully. As theories involve ideas about the relationships that exist among specific variables, so metatheories involve ideas about the relationships that exist among classes of variables.

For example, the second author was recently involved in a study regarding changes in human information processing following extended stress (Harris, Hancock, & Harris, 2005). In this study, we considered the relationship of such variables as spatial processing, mood, fatigue, and stress. A theory of the relationship of these variables can be situated within a metatheory involving large classes of variables: not just spatial processing, but all cognitive processing; not just mood, but all individual differences variables; not just task stress, but all aspects of a task, as well as the energetics of the environment. Situating such a study within a metatheory might help us, in future work, to consider additional important variables, and make connections to additional bodies of theory.

Why a Metatheory?

Certainly the question must arise, why bother with a metatheory? We focus here on two major reasons.

First, a metatheory can alert researchers to classes of variables that they might otherwise ignore. Every sub-discipline in any given era has its variables that enjoy more-favored and less-favored status, regardless of the actual importance of these variables for the phenomena under consideration. HFE is no exception to this rule. Thus, we find it interesting to note how frequently individual differences variables are excluded from research on human or team performance, even when individual differences variables are clearly implicated in such performance. Paying attention to a metatheory might help researchers include important variables that might otherwise be ignored. (In addition, as we present in the example below, a metatheory can alert one to potential *interactions* among variables.)

Second, a metatheory can alert researchers to other theories that might have bearing on the phenomenon being researched. For example, once a researcher recognizes that it is important to pay attention to certain individual differences variables, this should not just result in the researcher merely tacking an instrument to assess some such variable onto a research protocol for the sake of theoretical inclusiveness (the research equivalent of tokenism, we suppose). Such procedures just generate vague and imprecise research designs. Rather, researchers should take it upon themselves to learn something of the theories involving those variables that they have been prompted to consider including. This can result in more thoughtful research design and theory building.

Outlines of the Metatheory

Our metatheory of HFE describes any given achievement-related situation in terms of several major units (large super-classes of observables or constructs; see Figure 1). Our particular metatheory is oriented to situations involving an individual operator, even if that operator is acting within a team context. The major units of the metatheory are:

- *Task*: The task is comprised of that which is assigned to the personnel (see below). It should be noted that the "task," as we conceive of it here, is conceptual; what the personnel actually does is performance (see below).
- *Personnel*: The personnel is the human organism to whom the task is assigned.
- *Tools*: These are the instrumentalities through which the personnel effects his/her performance. Tools may

include, for example, a single hammer, or the multiple control panels of a starship.

- *Performance*: Performance is the behavior actually emitted by the personnel, in response to the assigned task.
- *Environment*: Environment involves the entire context in which the task is assigned, in which the personnel acts, and in which the personnel's behaviors are emitted. This includes the social environment, as well as any intelligent agents (e.g., AI constructs) involved in the situation.

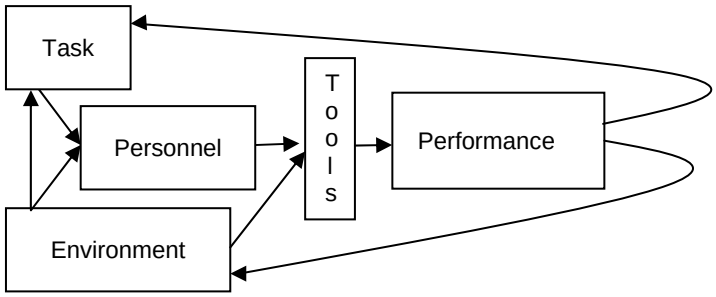


Figure 1. *Metatheory of human factors and ergonomics.*

A theory is not just a collection of variables, but is also a description of the relationships among those variables. Similarly, a metatheory is not just a collection of classes of variables, but is also a description of the relationships among those classes. In the metatheory that we have defined, the task influences the personnel, who, through her/his performance, may alter the nature of the task (even if only by accomplishing it and crossing it off the to-do list, as it were). The environment affects tools as well as the personnel, who, through performance, may in turn influence the environment. The environment may change the nature of the task, as well.

Some examples are in order concerning these relationships, and their feedback mechanisms. Changing circumstances in the environment may change the nature of the task "on the fly," as it were (e.g., the sudden appearance of a group of enemy airplanes really should change the nature of the task facing a military pilot). In terms of feedback, it is clear that, in many instances, performance introduces changes into the environment. (For example, the product is made, the signal is sent, the injury is treated, the enemy factory is destroyed, etc.) As another kind of feedback, performance may change the nature of the task. (For example, performance, as supervised by management, command, or an interactive AI agent, may prompt supervisors to change a task.)

Tasks include these components:

- instructions;
- statements regarding mandatory and optional objective(s) or outcome(s);
- statements regarding recommended or required procedures to obtain objectives;
- other information about the task (e.g., background, performance consequences), supplied to personnel.

Personnel include the following components (see Koltko-Rivera, 2005):

- perceived physical workload (subjective);
- mental workload experienced by the personnel;
- consciousness characteristics (e.g., current affect and surface cognition, situation awareness, situated cognition);
- physical aptitude (e.g., strength, immune system functioning);
- physical conditioning (e.g., training, prior experience);
- mental aptitude (e.g., intelligence);
- mental conditioning (e.g., training, prior experience);
- personality characteristics (e.g., traits, personality superstructure);
- motivation (e.g., need for achievement, motivation to learn, position on Maslow's hierarchy of needs);
- homeostatic and metabolic mechanisms;
- worldview (i.e., assumptions about reality's "is"s and "ought"s; values; Koltko-Rivera, 2004; Koltko-Rivera, Ganey, Dalton, & Hancock, 2004);
- acculturation (i.e., investment in a specific culture's behavioral expectations; Koltko-Rivera, Ganey, Dalton, & Hancock, 2004).

Environments include the following components (see Koltko-Rivera, 2005):

- objective physical workload imposed by the environment;
- visual/optical characteristics of the situation;
- auditory characteristics;
- haptic characteristics;
- olfactory characteristics;
- gustatory characteristics;
- kinesthetic characteristics (including acceleration/gravitation);
- electromagnetic/radiological characteristics;
- thermal characteristics;
- toxic characteristics;
- threat characteristics;
- nutritive characteristics;
- social characteristics (including the presence of intelligent agents);
- resources supplied to personnel.

Tools include the following components:

- objective physical workload imposed by use of the tools;
- visual/optical characteristics;
- auditory characteristics;
- haptic characteristics;
- olfactory characteristics;
- gustatory characteristics;
- kinesthetic characteristics associated with use of the tools;
- electromagnetic/radiological characteristics;
- thermal characteristics;

- toxic characteristics;
- threat characteristics.

Performance includes the following components:

- task-relevant behavior (e.g., action taken, as well as the reaction time, speed, and accuracy involved);
- task-irrelevant behavior.

An Example of Metatheory Use

Consider the situation of the applied researcher who is designing a research project to evaluate the advantages and disadvantages of alternative designs of a portion of a control room for a space station. This portion of the station involves communications control, which defines a set of tasks; the space station itself (along with its social dimensions and the presence of any AI constructs) defines the environment; the anticipated staffing of the space station, including the communications staff, define the personnel; expected tasks are set out by command; the communications equipment and control panels are the tools; actual performance is yet to be determined, but potential performance involves, for example, speed and accuracy of personnel response to commands.

In designing the research, the researcher must make choices of variables to include in the project, and have some sense of how these variables relate to one another. Taking the metatheory into consideration, the researcher might consider variables that might otherwise be ignored.

Within the domain of environment, of course it would be necessary to include the kinesthetic (microgravity) and electromagnetic/radiological aspects of the space station. However, taking the metatheory into consideration, one might consider an unusual interaction of variables that is specific to the space station situation. Does the microgravity environment of the space station influence the optical characteristics of the tools? After all, the human eye normally functions under conditions of gravity. Might the microgravity environment, through altering the shape of the human eye, affect the optical characteristics of a piece of equipment or a display, and thus performance in that environment?

Within the domain of personnel, it would be normal in an HFE project to include measures of subjective mental workload and situation awareness, as well as measures of physical and mental conditioning (e.g., training). However, taking the metatheory into consideration, the researcher might wish to consider worldview issues (Koltko-Rivera, 2004). In particular, the researcher might wish to consider worldview variables such as relation to group (also known as individualism-collectivism); in the close environment of a space station, not unlike a submarine, surely performance may be affected by whether the operator takes the position "one for all" (the collectivist approach) or "all for me" (the individualism position). Such differences in turn may imply design considerations.

After conducting the research, the results may carry implications for theory in the areas involving the variables chosen. Thus, in the research report, the researcher may not only comment on the implications of the research for the selection of control panels, but may have contributions to make regarding theories of how personnel worldview and environmental microgravity affect performance.

CONCLUSION

There are compelling reasons for HFE researchers to pay attention to theory in planning and reporting research. The metatheory that we have proposed will, we hope, facilitate this process.

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. Dr. Hancock's portion of this work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

REFERENCES

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Dubin, R. (1978). *Theory building* (rev. ed.). New York: Free Press.

Hancock, P. A., Weaver, J. L., & Parasuraman, R. (2002). *Sans* subjectivity—ergonomics is engineering. *Ergonomics, 45*, 991-994.

Harris, W. C., Hancock, P. A., & Harris, S. C. (2005). Information processing changes following extended stress. *Military Psychology, 17*, 115-128.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt Brace.

Koltko-Rivera, M. E. (2004). The psychology of worldviews. *Review of General Psychology, 8*, 3-58.

Koltko-Rivera, M. E. (2005). *A component-oriented metatheory of human behavior: A 'grand theory' justified and described.* Manuscript in preparation.

Koltko-Rivera, M. E., Ganey, H. C. N., Dalton, J., & Hancock, P. A. (2004). Worldview and acculturation as predictors of performance: Addressing these variables in human factors/ergonomics research. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 1223-1227). Santa Monica, CA: Human Factors and Ergonomics Society.

Koltko-Rivera, M. E., & Hancock, P. A. (2005). *A metatheory of human factors and ergonomics science.* Manuscript in preparation.

Kukla, A. (2001). *Methods of theoretical psychology*. Cambridge, MA: The MIT Press.

Slife, B. D., Reber, J. S., & Richardson, F. C. (Eds.). (2005). *Critical thinking about psychology: Hidden assumptions and plausible alternatives*. Washington, DC: American Psychological Association.

Slife, B. D., & Williams, R. N. (1995). *What's behind the research? Discovering hidden assumptions in the behavioral sciences*. Thousand Oaks, CA: Sage.

CONTACT INFORMATION

The first author may be reached by e-mail sent to mark@professionalservicesgroup.net. The second author may be reached by e-mail sent to phancock@pegasus.cc.ucf.edu.

On the Relationship between Workload and Performance

T. Oron-Gilad*, J.L. Szalma, S.C. Stafford, and P.A. Hancock

University of Central Florida

* Corresponding author

Tal Oron-Gilad
Institute for Simulation and Training & Department of Psychology
University of Central Florida
3100 Technology Pkwy, suite 337
Orlando, FL, 32826
USA

Tel: 407-823-0923
Fax: 407-823-0921
Email: torongil@mail.ucf.edu

Abstract

We studied the relationship between perceived workload and performance by evaluating the responses of professional police officers to different task demands inherent in two separate field training exercises. In the first of these exercises, designated as study one, seventy-one participating officers were required to complete a target shooting exercise as part of their regular training regimen. Results from this study confirmed the pattern of workload-performance associations and insensitivities as predicted by the extended-U model of stress effects. Study two was conducted in the same training environment but with significantly reduced task demands. The absence of change in either workload or performance in these latter conditions acted as further confirmation of the extended-U description while contradicting the competing inverted-U description relating stress and workload to performance.

Keywords: Workload; Field-Study; Police Officers; Performance Capacity, Workload Relationships.

On the Relationship between Workload and Performance

It is axiomatic that the output performance of any individual is a direct reflection of their momentary response capacity. But to what degree is this current level of response indicative of that individual's future performance? How do we know whether this particular moment's performance represents their absolute maximum response or conversely whether they are presently 'coasting' with a significant level of residual capacity left over with which to cope with any additional demands placed on them? We can and, since the time of Fredric Winslow Taylor certainly have, made inferences about future capacity predicated upon present response. However, are there better ways? Such concerns were among major imperatives for the growth of research in mental workload assessment; a topic now several decades old (see Moray, 1979; Hancock & Meshkati, 1988). There are numerous ways in which mental or cognitive workload might be assessed, independent of performance on the primary task of concern (Meshkati, Hancock, & Rahimi, 1989). One can either impose a secondary or loading task on the individual or take a suite of physiological measures in the search for diagnostic indices. However, the specific measure which has emerged largely because of its simplicity of administration and its face validity, is subjective response. There are numbers of measures of subjective response but perhaps the two most popular and currently predominant are the NASA Task Load Index (TLX: Hart & Staveland, 1988) and the Subjective Workload Assessment Technique (SWAT: Reid & Nygren, 1988).

For any measure of mental workload to be effective, it is important that it is diagnostic with respect to the near future performance on the task to hand (O'Donnell & Eggemeier, 1986). Workload measures can be highly diagnostic or they may be low on diagnosticity and more reflective of general situational demands (Wierwille & Eggemeier, 1993). If the various measures of workload always

slavishly followed performance[1] (i.e., a deterministic *association*; see Hancock, 1996; Parasuraman & Hancock, 2001; Yeh & Wickens, 1988) such measures would have little or no informational value except in circumstances where it was for some reason impossible to assess performance itself. Similarly, if there was a complete and persistent *dissociation* between a subjective measure of workload and performance, the former measure would have little practical utility. In some performance tasks such as vigilance, reports have recorded a direct and consistent association between performance and subjective mental workload (e.g., Warm, Dember, Gluckman, & Hancock, 1991; Warm, Dember, & Hancock, 1996; and see also Szalma et al., 2004). In these situations, increases in task difficulty simultaneously induce both performance decrement and an increase in subjective workload. In contrast, investigators examining multiple-task conditions (e.g., Yeh & Wickens, 1988) have reported a dissociation between perceived workload and performance. This latter observation indicates that workload is not simply a ubiquitous reflection of task difficulty. To date, there has been very limited theoretical or empirical research to articulate the factors that influence this performance-workload relationship. Given that different task domains produce different performance-workload relations, a central question emerges: what information is provided by, and what factors drive subjective measures of workload that are not intrinsic to performance measures. The present study examined the pattern of this performance-workload link but rather than use a laboratory setting, we evaluated this relationship in the context of a real-world operational setting to which any such understanding will eventually have to be applied.

In an attempt to distinguish the formal link between workload and performance, Yeh and Wickens (1988) provided an approach using a framework that is grounded in attentional resource theory (Kahneman, 1973; Navon & Gopher, 1927; Wickens, 1980). They proposed that dissociation between performance and workload occurs under a number of circumstances that include the following: i) when greater resources are invested to improve resource-limited tasks (cf., Norman & Bobrow, 1975); ii) if

[1] Although such primary task performance is often considered the major reflection of mental workload (see Hancock & Meshkati, 1988), here we take it as the criterion measure against which other measures are necessarily compared.

demands on working memory are increased by time-sharing; and iii) when performance is sensitive to some sub-task element while subjective measures reflect more global demands. Dissociations can be examined from a theoretical perspective and once better understood some situations for their occurrence can be anticipated (Derrick, 1988). Dissociations can also occur when greater resource investment is induced through increasing motivational incentives (e.g., Vidulich & Wickens, 1986). In addition to association and dissociation, insensitivities have also been observed. Articulated by Hancock (1996; see also Parasuraman & Hancock, 2001), insensitivities represents the cases in which either the primary task performance or the subjective workload changes but the other does not [the cases that Hancock (1996) referred to as 'insensitivities' were considered forms of 'dissociation' by Yeh and Wickens (1988)] but the full pattern of possible relationships indicates that dissociations have to be distinguished from insensitivities (see Table 1). Hancock (1996), argued that as well as dissociations, insensitivities are also diagnostic. Thus, a case in which performance is stable among conditions but subjective workload increases indicate a tradeoff in which the individual maintains performance only by exerting more effort. By contrast, when performance changes but workload does not, the individual shows that at that moment they are oblivious to their own task response level.

<Insert Table 1 about here>

Some researchers have argued that workload is a one-dimensional concept (e.g, Wierwille & Cascali, 1983; Colle & Reid, 1999; Hendy, Hamilton, & Landry, 1993). Others have argued that workload is a multi-dimensional concept (e.g., Derrick, 1998; Gopher & Donchin, 1986; Vidulich & Tsang, 1986). Among the latter, there is much less agreement about the individual dimensions and their specific contribution to the perception of workload or to the workload-performance relationship (see Nygren, 1991; Reid & Nygren, 1988; and recently Rubio, Diaz, Martin, & Puente, 2004). Resolution of

these issues is critical if subjective reports of workload are to be meaningfully related to the objective performance parameters (see Annett, 2002; Natsoulas, 1967).

Perceived workload can also reflect of the level of stress that the individual experiences while performing a task (Hancock & Desmond, 2001). Tasks themselves, as Hancock and Warm (1989) have noted, are often the proximal source of stress. The Hancock and Warm model of adaptation under stress predicts that increased perceived workload precedes any performance change and therefore a systematic pattern of workload-performance insensitivities and dissociations are expected as long as the individual remains within the 'psychological zone of maximal adaptability'. This sequence is shown in Figure 1 and is linked to a more formal description in Table 2. Contemporary models of human performance are largely founded on the findings of the experimental laboratory (Hancock, 1989). In general, such models often neglect the whole spectrum of energetic characteristics of response which are influential in real worlds operations (but see Hockey, 1986). It is precisely under these conditions that information concerning operator and system response becomes most valuable. Practical investigations such as the one presented here, are rarely explored and more data on how the psychological models work in a real-world settings is crucially needed.

<Insert Figure 1 and Table 2 about here >

PERFORMANCE EXERCISES

The particular circumstances investigated here are two different police shooting field-training exercises which were part of the routine training program of police officers in the South-Eastern part of the United-States. These training procedures on shooting proficiency are intended to both improve and maintain officers' shooting skills. They are conducted approximately two times per year. The first night shooting exercise (Study I) consisted of four shooting and maneuverability tasks that increased in

difficulty as the training progressed. The second study was a qualifying training exercise on a novel weapon (Study II) consisted of three courses. Specifically, officers were familiar with the general procedure that the exercises followed, with the instructing officer and his assistants, and with the location of the exercise at the departments' outdoor shooting range. Both exercises were not intended to be intimidating. Those who failed to meet the satisfying criteria for passing the training exercise were called for additional training until they reached a satisfactory level of proficiency.

STUDY I – NIGHT SHOOTING EXERCISE

Night training exercises are not obligatory in the police training regimen in the Southeastern United States. The particular Police Department with whom we have been working has voluntarily been conducting night shooting exercises on a yearly basis as part of their officers training program. In this case, it is the training officer's responsibility to design the exercise and the shooting task manipulations based on his experience and understanding. As a discretionary exercise, there are no official instructions or qualifying (pass/fail) requirements.

Experimental Method

Experimental Participants. All police officers in this particular Department were required to participate in the night shooting exercise as part of their continuing training. However, participation in our research component was voluntary and each individual officer was free to decide whether or not to participate in our procedure. Of the 91 officers who completed the exercise, 71 (78%) volunteered to participate in our research. These included 10 women and 61 men, who had a mean age of 37 years (range 22-56 years) and with an average of 11 years of police experience (range 1-32 years). In some cases officers did not complete the all of the questionnaires between shooting tasks as requested. This

left a total of 62 participants who completed the whole data series and it was these data which were subjected to analysis.

Firearms. Each officer used a SIG (Schweizerische Industrie Gesellschaft™) Sauer p226 9mm handgun. All officers used the same standard issue duty belt and equipment. All officers also carried a standard Department issued Streamlight Stinger flashlight which was 19 centimeters long and weighing 284 grams which they used in three of the four tasks. The only task which they were not required to hold the flashlight while shooting was the Warmup Task (WT; and all tasks are described below). All officers were required to wear sound attenuating hearing protection (external muffs or internal plugs), body armor, and clear eye protection during each shooting task.

Task Design. The night exercise included four different shooting tasks. Those were: i) the 'Warmup Task' (WT), ii) the 'Basic Flashlight Task' (FT), iii) the 'Barrel Task' (BT) and iv) the 'Metal Task' (MT). Three of the tasks (WT, FT, and BT) were held in the same shooting environment. The MT was a novel task and had slightly different requirements as we detail below. The tasks increased in difficulty as the training session progressed. The 'Barrel Task' and the 'Metal Task' were the most complex of the four. The tasks were pre-designed by the training officers and so there was no opportunity for any experimental control directly. All officers were familiar with the WT and with the use of a flashlight while shooting (FT) from their previous training exercises. The BT and the MT were introduced for the first time in this exercise. However, as mentioned previously BT was a more demanding version of the FT and as such more difficult than the FT but not entirely novel. The only research-intervention permitted was that the order of the two more complex tasks (BT and MT) varied among participants, hence the order of the tasks was either WT, FT, BT, and MT, or WT, FT, MT, and BT. However in keeping with the non-interventionist nature of this study we had no control as to who was assigned to either of the orders and therefore we cannot claim homogeneity of the number of participants in each group or homogeneity of variance of performance ability.

Warm-Up Task (WT) – 'Baseline' Shooting Skill. The first two tasks (WT and FT) were group exercises where teams of officers participated at the same time under the overall direction of the training officer. WT was a basic draw and shoot task in which each officer was required to fire a total of 16 rounds from 4 different distances (1.5, 4.5, 6.5, and 10.5 meters, respectively) at a standard paper target which presented a white silhouette of a human figure. All rounds were fired in very low ambient light conditions in which the white silhouette of the target was barely visible. All officers were operating under the command of the instructor and were to draw their weapon and engage the target upon the verbal command 'gun'. Officers were instructed not to move while engaging the target. For each task officers who performed below the level of instructor expectations were required to repeat the task on the following training day (for the purpose of this study, data was recorded for each officer once on the first day of training). Performance on this task was evaluated by the percentage of hits out of the sixteen shots fired. A hit is a shot placement that is consistent with a wound which is expected to kill or incapacitate an opponent, a miss is a shot placement which missed the silhouette of the target or was inconsistent with a mortal or incapacitating wound (e.g., a hit in the arm of the target figure).

Flashlight Task (FT) – Basic Flashlight Control and Movement Exercise. FT was a twenty-four round shooting exercise with reloads, while holding the flashlight and moving (i.e., rotating the body position to face 90 degrees or 180 degrees away from the target and then engage). Some portions of the task required officers to fire their weapons in very low ambient illumination conditions, whereas at other points officers were required to illuminate targets with their flashlights. In addition to shooting with a flashlight, officers had to reload their weapons including two plastic 'dummy' rounds into two of their three magazines. This provided situations where they were to reload the weapon before the last round in the magazine was expended. The dummy rounds imposed an additional task demand by jamming the weapon and requiring the officer to clear the weapon of the dummy round and continue to engage the

target with appropriate number of rounds. All rounds in this task were fired from a fixed distance of 6.5 meters. Performance on this task was evaluated by the percentage of hits out of the 24 shots fired.

The 'Barrel' Task (BT) – Movement and Taking Cover while Shooting Exercise. Unlike the first two tasks, which were group exercises, this task was performed by each officer separately. In this task, the officer was required to shoot as quickly as possible shoot at a sequence of five targets, arrayed horizontally and equally distributed over approximately 27 meters. The officer began the task by facing away from the first of the 5 targets with a loaded weapon. On the instructors command, he or she engaged the five targets at different ranges with a specific number of rounds per target. After engaging each target, the officer ran to the next shooting location to engage the next target. On targets two and four, the officer took cover behind a plastic barrel and illuminated the target with a flashlight prior to engaging it. Performance on this task was evaluated by the percentage of hits out of the 15 shots fired.

The 'Metal' Task (MT) – Feedback Shooting Exercise. Like the BT, this task was performed separately by each officer. The task was conducted in a different area in the range illuminated by the strobes of a police cruiser. The officers were instructed to complete the task as quickly as possible. The task began with the officer facing away from a 4.5 meters wide wooden wall with a 60 cm by 91 cm opening in its center. When the pro-timer sounded with a loud audible beep the officer had to draw both flashlight (to illuminate the target) and gun and look through the opening. Seven meters away were 6 metal targets posted on metal poles. They were spatially separated into 2 groups of 3 and each group included a white target, a red target, and a blue target. As soon as the officer pointed the weapon toward the targets, the instructor called out a color. The officer had to shoot (and hit) each metal target in that color twice consecutively (4 shots altogether). Once completed, the officer was instructed to pass the wooden wall, and stop at the second firing line (a total movement of approximately 14 meters) which was 9 meters away from seconds set of 8 metal targets separated into 2 groups of 4. Each group had a

white, black target, yellow target, and red target. Similarly, the instructor called out a color and the officer had to hit each target with that particular color for 2 consecutive rounds. The timer was stopped when the officer completed hitting the targets. Performance on this task was evaluated by the duration of the individual's performance time and by the percentage of hits out of the total shots fired (note that in this task the number of hits was constant (8) while the number of possible shots varied among officers). Contrary to the BT, the MT provided feedback after each shot as a result of the sound of the bullet hitting the metal target (the hearing protection provided attenuation of the noise but did not completely mask the sound of the bullet hitting the target). Also, the metal targets differed in shape from the paper targets and officers were not familiar with shooting at metal targets. Also in the MT officers were required to aim at one of three (for the shorter distance) or four (for the longer distance) possible targets. In contrast, in the BT, officers shot 5 different targets sequentially with only one possible target at a time.

Performance Assessment. There are no State enforced rules as to passing this shooting exercise. The State requirements specifically address shooting performance during the day in a stationary position (e.g., the qualifying courses of Study II). Performance levels on the night training are therefore left to the judgment of the training officer. Individuals who the training officers thought did not perform adequately on any one of the tasks was required to return to the shooting range on successive days until the training officers were satisfied with their level of performance.

The Time Estimation Secondary Task. During the BT and the MT officers were required to estimate the duration of the task following their performance. Zakay and his colleagues (Zakay & Block, 2004) have demonstrated that the prospective duration estimation (when a person is aware that a duration judgment will be made at the end of the task) can serve as a secondary task for workload measurement (and see Block, Zakay, & Hancock, 1999). Zakay (1998) has suggested that

underestimation is more likely to occur when time estimation is not the primary task at hand. Allocating fewer attentional resources to temporal information processing causes fewer time signals to accumulate, resulting in a decrease in the estimated duration.

Subjective Workload Assessment. Mental workload was assessed using the Raw NASA Task Load Index (RTLX) scores, an un-weighted average of the subscale values. In the original version of the TLX (Hart & Staveland, 1988), paired comparisons were used to derive weights for the six subscales of the TLX. However, as Byers, Bittner, and Hill (1989) have shown and as also discussed in length in Nygren (1991), RTLX scores can provide an even better account of the workload experienced by the participant than traditional weighted TLX scores. The RTLX is composed of six sources of workload which are mental demand, physical demand, temporal demand, frustration, effort, and self-rated performance. Following each task, officers rated their perceived workload on these individual sources each on a scale from 0-100. Global RTLX estimates were derived by the calculating the average of the subscale values.

Experimental Procedure. The exercise occurred in January 2004 at an outdoor police shooting range in the Southeastern part of the United States. The average temperature at this period of time was approximately 10° Celsius with no rain. The exercise was performed in the dark (after 7PM), and the total session time varied from 60 to 120 minutes depending on the number of officers in the respective session. Officers were trained in groups of 6 to 12 at a time. Prior to the beginning of the exercise the officers were pre-briefed by the training officer on safety precautions and on task composition. The pre-briefing was held in a well lit lecture room which was part of the shooting range facility. At the end of the pre-briefing the officers were briefed on the research and its purpose and then asked if they would volunteer to participate. The experimenter then presented the RTLX questionnaire and briefly explained its components. Officers were asked to complete the RTLX at the end of each task. The forms were

distributed and completed in a lighted area behind the shooting range where the officers loaded their weapons throughout the exercise. Additionally, the officers were told that they would be asked by the experimenter to provide an estimate of the task length (prospective duration estimation) in minutes and seconds upon completion of the BT and the MT. A paper based version of the RTLX questionnaire was filled after completion of each task and time estimates were collected upon completion of the BT and the MT.

Results

Overall Performance. Shooting performance was assessed separately for each task by calculating the percentage of hits from the total shots made (the definition of a hit is based on the Police Department's criteria for a hit and not on merely hitting the target). This measure is the simplest and most direct reflection of performance. However, within each task the shots were made from different ranges, causing shooting accuracy and difficulty to vary even within a single task but this influence could not be controlled. However, accuracy does provide an overall indication of how well each officer performed. For the BT and the MT, which were performed individually by each officer, task duration data was collected and therefore a speed-accuracy tradeoff was able to be examined.

Accuracy. The descriptive statistics and confidence intervals for hit percentages are presented in Table 3. A repeated measures ANOVA on hit percentage showed significant effects ($F(3,61)=13.54$, $p<.0001$). Based on the confidence intervals, we conducted three, 2-tailed *t*-tests on the respective differences. Only the differences between FT and BT proved significant ($t(61)=2.69$, $p<.009$). The other comparisons between FT and MT ($t(61)=1.18$, $p>.05$, Cohen's $d=.16$), and between BT and MT ($t(61)=1.26$, $p>.05$, Cohen's $d=-.18$) proved insignificant. No order effects were found for changing the sequencing of the tasks between the MT and the BT ($p>.05$). Again, we must note that we had no control

over the homogeneity of groups. Furthermore, years of service were not a significant moderator of performance accuracy. Significant correlations were observed between the WT and the FT (r=.58, $p<.01$), between the WT and the BT (r=.47, $p<.01$), and between the FT and the BT (r=.45, $p<.01$). MT did not correlate significantly with any one of the other tasks.

<Insert Table 3 about here>

Speed-Accuracy Tradeoffs. Duration for the BT was 59 seconds (SD=13 seconds) and for the MT 57 seconds (SD=19 seconds). For each task the participants shot 15, and 8 shots accordingly. For the BT there was no correlation between the duration of the task and accuracy (hit percentage) (B=-.19, SD=.19, β=-.12, t(70)=-.99, p>.3). For the MT there was a significant correlation between the duration of the task and hit percentage (B=-.28, SD=.92, β=.346, t(70)=-3.0, p<.003) indicating that the better performers also performed the task faster.

Workload Estimates. Perceived global workload was derived from the (un-weighted) average of the six sub-scales. Descriptive statistics and confidence intervals for hit percentage are presented in Table 4. A repeated measures ANOVA on hit percentage was significant (F(3,61)=56.27, p<.0001). Based on the confidence intervals, we conducted three 2-tailed t-tests on the difference between FT and BT (t(65)=-8.10, p<.0001, Cohen's d=-.70), FT and MT (t(66)=-5.95, p<.001, Cohen's d=-.49), and BT and MT (t(66)=3.28, p<.002 Cohen's d=.24), those revealed a significant differences in workload ratings between the MT and the BT and between the MT and the FT. No order effects were found for changing the order of tasks between the MT and the BT. Years of service again were not a significant moderator of workload ratings. Table 5 provides the descriptive statistics of the NASA-TLX subscales for the four tasks.

<Insert Table 4 and 5about here>

Association and Insensitivity. The pattern of workload-performance relationship that has been identified from the performance and global workload estimates is one of both associations and insensitivities. Associations, in which increased task difficulty was associated with decreased performance and increased workload, were observed for the WT, FT, and BT tasks respectively. Performance insensitivities, where workload increased but there was no significant change in performance, were observed between the FT and the MT and between the BT and the MT.

Separate Task Analyses for the BT and the MT. To investigate the source of the insensitivities the two more complex tasks, the MT and the BT were examined in more detail. A hierarchical regression model of workload and performance was computed. This model was not intended for performance prediction but rather for explanation regarding which workload dimensions affect (or were affected by) performance and the strength of these effects (see Cohen, Cohen, West, & Aiken, 2004). For the BT, a hierarchical, linear regression was performed on shooting performance on the BT with *shooting skill* (defined here as shooting performance in the baseline shooting task (WT)) entered first, followed by each *TLX subscale.* This was done so that regressions could be compared to determine whether shooting performance in BT was mediated by the initial shooting skills while examining the relationship between the six individual TLX subscales and *shooting performance*. For BT, the only model where a subscale added significantly to the prediction of the *shooting performance* was the regression model with the *Mental* subscale ($B_{\text{shooting skill}}$=.48 (SE(B)=.11), β=.44, p<.0001, R^2=.22 and $B_{\text{mental subscale}}$=-.25 (SE(B)=.10), β=-.26, p<.05, ΔR^2=.07 for Step 2 (ps<.05)). For the MT a similar analysis was conducted. Unlike BT, overall performance in MT did not correlate with the WT, indicating that *shooting skill* did not influence *performance* in the MT. In order to examine the relationship between the six individual TLX subscales and MT *shooting performance*, a series of hierarchical linear regressions were performed

on shooting performance with the individual TLX subscales. Only the models in which *Frustration* and *Own-Performance* were entered showed a significant predictive power. The most predictive model was one that combined both subscales (($B_{\text{frustration subscalel}}$=.18 (SE(B)=.06), β=-.32, p<.005, R^2=.14 and $B_{\text{performance subscale}}$=-.23 (SE(B)=.07), β=-.34, p<.005, ΔR^2=.12 for Step 2 (ps<.005)). Note that there was no significant correlation between these two subscales, r=.050.

Time Estimation. Prospective time estimations were made for the BT and the MT. The mean *actual* time taken was 59 seconds (SD=13) for the BT, and 57 seconds (SD=19) for the MT. The mean *estimated* time was 64 seconds (SD=44) for the BT, and 66 seconds (SD=51) for the MT. Actual time and estimated time scores were used to compute two derived time estimation measures; i) the duration judgment ratio (DJR; Block, Zakay, & Hancock, 1999), which is the ratio of the estimated time and the actual time in minutes and seconds, expressed as a percentage, and ii) the absolute error (AE) which represents the discrepancy between the estimated time and the actual time (also in percentages). DJR for the BT was 109 (SD=68) and for the MT was 116 (SD=70). For the BT, 57.5% of the officers underestimated the duration of the task (i.e., thought that the task was shorter than it actually was). For the MT, 59.2 % of the officers underestimated the duration of the task. The inter-participant variability was large and the absolute error measure indicated that on average, time estimations were quite distant from the actual time (44% for the BT and 49% for the MT, respectively). Underestimation of time was expected considering that time estimation was not the primary task (Zakay, 1998). The correlation between the DJR of the two tasks was significant (r = .456, p<.001), and respectively, the correlation between the absolute error (AE) of both task was also significant (r =.288, p<.05). Thus, officers were relatively consistent in the estimation of time across tasks. Correlations between workload and time perceptions, as well as correlations between shooting performance and time perceptions were examined.

However, neither global workload (nor any of the workload subscales) nor performance were significantly correlated with DJR nor with the absolute error for either task.

Summary for Study I

Associations were observed for the WT, FT, and BT while performance insensitivities were observed between the FT and the MT and between the MT and the BT. The increase in task difficulty and maneuverability for the WT, FT, and BT was reflected by degradation in hit rate and a consistent increase in workload. However, the significant correlations among these tasks may indicate that these three tasks, based on shooting a single paper target per trial were to some extent 'more of the same'. That is these three tasks required similar skills. In spite of the novelty of the BT, which was the most complex of the three, it was closer to what the officers had expected and similar to what they routinely do on the range, 'just slightly more difficult' and more physically exerting (see Table 5). This was not the case for the MT where variability in performance was due to other unrelated factors. This difference may be due to the metal targets themselves as distinct from the paper targets but also may be due to the immediate feedback provided by hitting the metal targets. Alternatively, it may be a result of the uncertainty regarding which target color was to be selected by the training instructor, uncertainty which was not present in the other three tasks.

STUDY II – QUALIFYING COURSE FOR A NOVEL WEAPON

Study I examined four different tasks that varied in task difficulty. However, it can be argued that under variable conditions it might be quite difficult to determine the source of insensitivity and dissociation of subjective workload from performance. we therefore examined the repetitions of one single task over the course of a training exercise. In contrast to the first study, in this procedure, the

officers had to follow a specific State-enforced qualifying procedure in order to be authorized to use a the newly introduced weapon.

Experimental Method

Experimental Participants. All officers were required to participate in the shooting exercise as part of their continued training. However, participation in the research was again voluntary. Of the 25 officers who participated in the training exercise, 22 officers (88%), 1 woman and 21 men, volunteered to participate in the research component.

Firearms and Holsters. In this exercise the officers were introduced to new weapons and new holsters. The new weapon was the SIG P226 (.40 caliber Smith & Wesson (S&W)) handgun. This weapon replaced the SIG P226 (9mm) handgun. The difference between the weapons being in the unloaded weight and magazine capacity. The .40 S&W has a slightly larger barrel and its unloaded weight is 65 grams heavier than the 9mm. Additionally, the .40 caliber bullet is slightly larger than the 9mm bullet. Hence, the .40 S&W holds 12 rounds in the magazine while the 9mm holds 15. Other than the slight increase in weight and the difference in magazine capacity, the weapons share the same physical outer dimensions, including grip size and trigger pull. The new holster locks the SIG P226 .40 S&W in place by a spring loaded lever that pushes the weapon forward into a block that engages the ejection port on the weapon. The holster that was used with the P226 (9mm) handgun was similar to the new holster except it used a different weapon release technique. All officers were required to wear sound attenuating hearing protection (external muffs or internal plugs), body armor, and clear eye protection during each shooting task.

Task Design. This exercise included 3 qualifying courses; one at the beginning of the exercise, the second after the completion of the first practice session, and the third after the completion of the

second practice session. Practice sessions varied in composition and were approximately of the same length (45 minutes).

Performance Assessment. Each qualifying course consisted of 13 rounds. Passing criteria to qualify was 9 hits. A hit (similar to Study I) was a shot placement that is consistent with a wound which is expected to kill or incapacitate an opponent. No performance measures were taken during the practice sessions.

Workload Assessment. Mental workload was assessed using the Raw NASA Task Load Index (RTLX), which provided an un-weighted average of the subscale values. Following each task, officers rated their perceived workload on each of these individual sources on a scale from 0-100. Global RTLX estimates were derived by the calculating the average of the subscale values.

Cognitive Performance Assessment (ANAM Readiness Evaluation System - ARES). ARES is a cognitive testing system designed to run a customizable batteries of tests. ARES Infantry version is a brief on-line assessment of the ability to sustain levels of concentration, working memory, and mental efficiency for operational commanders in field settings. The battery was presented on handheld computers (PDA's) which used the Palm OS version 3.5. For the purpose of the current study the 'Target Reaction Time' component was administered. This component of the battery consisted of two tasks: i) Target Detection, which is a simple response time task where the only "foe" targets appear on the screen and the operator is required to respond to the appearance of the "foe" targets; ii) Target Recognition, a go, no-go response time task where both "foe" and "friendly" targets appear and the operator should only respond to "foe" targets.

Experimental Procedure. The exercise took place during daylight hours (between 7AM and 2PM) at an outdoor police shooting range in the Southeastern part of the United States. Officers were trained in groups of 6 to 12 at a time and the duration of the exercise varied from 105 minutes to 120

minutes depending on the number of officers in the session. The average temperature was 33° Celsius (92° Fahrenheit) with humidity of 79%. After each qualifying course, the Officers completed a paper based version of the RTLX and ran the ARES Battery on a designated PDA. Due to procedural restrictions, the third RTLX was not administered.

Results

Qualifying Course Performance. The passing Criterion was 9 hits out of 13 (69%). During the first qualifying course 8 officers failed to pass and the average performance was 75% (SD=18%). During the second qualifying course 5 officers failed to pass (1 of them had already passed the first course) and the average performance rate was 79% (SD=17%). During the third qualifying all officers passed and the average performance score was 84% (SD=11%). Performance in the second qualifying course correlated significantly with performance in the third qualifying course (Pearson two-tailed, r=.63, p<.01). There were no significant differences in performance between the first and the second qualifying courses (p>.05, Cohen's d= -.27) or between the second and the third qualifying course (p>.05, Cohen's d= -.28), but there was a significant difference in performance between the first and the third qualifying course (t(20)=-.24, p<.027, Cohen's d= -.55). Note that a negative effect size indicates low-performance in the later qualifying course.

Workload Estimates. Perceived global workload was derived from the (un-weighted) average of the six sub-scales. There were no significant differences between the two measurements (M=40, SD=15 and M=39, SD=14 accordingly). There were also no significant differences between the subscales.

ARES Battery. For the Target Detection Task, response accuracy was 100% at all times and for all participants. Hence, none of the participant missed any of the responses. Average response times for the first, second and third time that the battery was administered were 298 ms (SD=74ms), 270 msc

(SD=46 ms), and 283 ms (SD=60ms) respectively. There were no significant differences between response times, however ω^2=0.08 a medium effect size, indicating that we should not assume the response times are statistically equivalent. For the Target Recognition Task, the average response accuracy was 92% (SD=10%), 93% (SD=9%), and 89% (SD=19%), respectively, and response times were 419ms (SD=17ms), 388ms (SD=11ms), and 393ms (SD=15ms) respectively. None of these differences were statistically significant and no speed-accuracy tradeoffs were identified. For response times ω^2=0.04 which is a medium effect size, indicating that we should not assume the response times are statistically equivalent. For accuracy ω^2=0.01 a small effect size, indicating that we can assume that accuracy scores are statistically equivalent.

Summary for Study II

These results provide an example of a situation where workload is providing little if any diagnostic value beyond the objective performance information. This is an example of what we have previously defined as the control situation. In terms of the Hancock and Warm (1989) model this is a case where task demands and overall stress never exceed the comfort range (see Figure 1 and Table 2) and this was confirmed by the general attitude of the officers involved. The ARES battery has been shown to be diagnostic under extreme stress conditions (See Harris & Hancock, 2005). However, it is not sensitive enough to capture the changes in operator state within the comfort range.

DISCUSSION

As we indicated in our initial introduction, the relationship between workload and performance is necessarily complex since the absolutes of total insensitivity or unvarying association are neither practically useful nor theoretically valuable. The value of workload is reflected in the time-varying

patterns of association, dissociation and insensitivities expressed as a function of dynamic change in task demand. Here we have postulated that the Hancock and Warm (1989) model of stress and performance can provide the principle basis for this pattern of workload-performance relationships. Specifically, the Hancock and Warm (1989) model predicts that increases in perceived workload will precede the onset of performance decrement. Therefore workload-performance insensitivity is expected given that: (a) the individual remains below the threshold of the 'psychological zone of maximal adaptability', and (b) the tasks gradually increase in complexity requiring the individual to exert more effort to maintain a stable performance level. In Study I we confirmed this pattern the demonstration that performance did not vary significantly between FT and MT and between MT and BT but that perceived workload did. This indicates that, in terms of the Hancock and Warm (1989) model, the officers were still within the threshold of the comfort zone (see (c) in Figure 1) where performance insensitivities are predicted to occur. Furthermore, only under extreme conditions do we expect time estimates to be predictive of the stress state of the individual (see Hancock & Weaver, 2005). This finding reinforces the 'shoulders' morphology of the extended-U relationship described by Hancock and Warm (1989) and is in sharp contrast to the popular but flawed, inverted-U description (Yerkes & Dodson, 1908; and see Hancock & Ganey, 2003). Such findings attest to the adaptability of the human operator; even though the performed tasks were complex, and required coordinated mental and physical work under time pressure in a sub-optimal environmental condition.

The insensitivities observed in Study I can be examined from the perspective of the multidimensionality of workload as reflected in the NASA-TLX subscales. For this purpose we investigated in more detail the insensitivity in performance that we found for the two more complex tasks MT and BT. This analysis showed that some dimensions of the overall workload scale contributed more than others to the workload assessments with shooting performance. For the MT, the self-rating

subscales were the better markers for objective performance correlation. Specifically, the *Frustration* and *Own Performance* subscales were the markers for performance for the MT. In contrast, for the BT, the *Mental* subscale combined with *basic shooting skill* were markers. The finding that for the BT only one subscale of the TLX was significantly correlated to performance and for the MT only two subscale correlated with performance presents an interesting quandary. Given the physical nature of the task, which involved running and taking cover while engaging targets, one might expect that the *Effort* and *Physical* demand subscales would exert important influences. Furthermore, researchers often tend to underestimate the ability of individuals to estimate their own performance (see Johnson, 2004) however in the MT underestimation would have been unjustified since the own-performance estimations accounted for a significant portion of the objective performance. Possibly, the immediate feedback provided in the MT made it possible for officers to estimate their performance and to attempt to improve it throughout the task (and see Newell, 1980). On the other hand, this immediate feedback may have also generated more frustration directly related to objective performance. For the BT, the finding that *Mental* demand accounted for such a significant proportion of variance in shooting performance indicated that that officers who expended the most mental effort were the poorest performers. This represents a workload-performance association (see Hancock, 1996). Although it is impossible to determine from the data in study I what was the cause of the differences between the MT and the three other tasks, it is clear that providing officers with tasks other than traditional ‘stand and fire’ shooting range tasks require different skill sets and thereby has the potential to enhance training outcomes. One likely reason that officers remained within the thresholds of the Hancock and Warm (1989) comfort zone is that training exercises like the ones presented here were not intended to physically challenge the participants. This is reinforced in Study II.

The results of the Study I underscore the importance of examining the subscales of the TLX. It has been argued that the subscales are not particularly useful (e.g., Nygren, 1991), but the results obtained here confirm that the TLX subscales have diagnostic value, an important criterion for any workload measure (O'Donnell & Eggemeier, 1986). Indeed, the pattern of association, dissociations, and insensitivities in the performance-workload relation can be diagnostic with respect to the differential effects of task characteristics on operators (see Hancock, 1996). To make full use of such information requires a better understanding of the forces that influence when such associations/dissociations and insensitivities occur. Empirical efforts to articulate such effects have been relatively rare and the present study was direct to systematically examine the performance-workload link using a real-world task where the pattern of associations and insensitivities obtained can have direct implications for actual training and job performance.

PRACTICAL IMPLICATIONS

The current environment in which law enforcement operate is one in which there is considerable workload and the price of failure with shooting can be fatal. To date, no studies of the performance workload relation have been done with police officers engaged in shooting tasks. This study is an initial exploration of this issue, and results indicate that measurement of performance and workload showed a pattern of association and insensitivity that is diagnostic. This indicates the need to expand police training with firearms beyond the traditional 'stand and shoot' procedures. Designing tasks that impose additional 'real-world' demand (e.g., holding a flashlight, running, aiming at more than one possible target at a time) may have beneficial effects in improving performance by forcing officers to practice shooting skills 'in context'. However, further research is needed to identify the additional task demands to be added in order to optimize training outcomes.

REFERENCES

Annett, J. (2002). Subjective rating scales: Science or art? *Ergonomics, 45*(14), 966-987.

Block, R.A., Zakay, D., & Hancock, P.A. (1999). Developmental changes in human duration judgments: A meta-analytic review. *Developmental Review, 19,* 183-211.

Byers, J.C., Bittner, A.C., & Hill, S.G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In: A. Mital (Ed.), *Advances in industrial ergonomics and safety, I* (pp 481-485). London: Taylor & Francis.

Cohen J., Cohen P., West, S.G., & Aiken, L.S. (2004). *Multiple regression/Correlation analysis for the behavioral sciences*. Third edition, Mahwah: New Jersey: Erlbaum and Associates.

Colle, H.A. & Reid, G.B. (1999). Double trade-off curves with different cognitive processing combinations: Testing the cancellation axiom of mental workload measurement theory, *Human Factors, 41*, 135-150.

Derrick, W.L. (1988). Dimensions of operator workload. *Human Factors, 30*, 95-110.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In: K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.), *Handbook of human performance. Vol. 2. Cognitive processes and performance* (pp. 41.1-41.49). New-York: Wiley.

Hart, S. G,. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P.A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Hancock, P.A. (1989). The effect of performance failure and task demand on the perception of mental workload. *Applied Ergonomics, 20,* 197-205.

Hancock, P.A. (1996). Effects of control order, augmented feedback, input device and practice on tracking performance and perceived workload, *Ergonomics, 39,* 1146-1162.

Hancock, P.A. & Desmond, P.A. (2001) (Eds.). *Stress, workload, and fatigue.* Mahwah: Lawrence Erlbaum Associates.

Hancock, P.A., & Ganey, H.C.N. (2003). From the inverted-U to the extended-U: The evolution of a law of psychology. *Human Performance in Extreme Environments, 7* (1), 5-14.

Hancock, P.A. & Meshkati, N. (1988) (Eds.). *Human mental workload*. Amsterdam: North-Holland.

Hancock, P.A. & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519-537.

Hancock, P.A., & Weaver, J.L. (2005). Temporal distortions under extreme stress. *Theoretical Issues in Ergonomic Science, 6*(2), 193-211.

Harris, W.C., & Hancock, P.A. (2005). Cognitive performance under high stress and high threat conditions. In preparation.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? *Human Factors*, 35(4), 579-601.

Hockey, G.R.J. (1986). A state control theory of adaptation and individual differences in stress management. In G.R.J. Hockey, A.W.K. Gaillard, and M.G.H. Coles (Eds.), *Energetics and human information processing* (pp. 285-298). Dordrecht, The Netherlands: Martinus Nijhoff Publishers.

Johnson, A. (2004). We learn from our mistakes – don't we? *Ergonomics in Design, 12*, 24-27.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Natsoulas, T. (1967). What are perceptual reports about? *Psychological Bulletin, 67* (4), 249-272.

Navon, D., & Gopher, D. (1979), On the economy of the human processing system, *Psychological Review, 86*, 214-253.

Norman, D.A., & Bobrow, D.B. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

Newell, A. (1980). The knowledge level. *AI Magazine*, 2(2), 1-20 (Presidential Address, American Association for Artificial Intelligence, AAA I80, Stanford, CA, August 1980).

Nygren, T.E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33*, 17-33.

Neville, M. (1979) (Ed.). *Mental workload*. New York: Plenum Press.

Meshkati, N., Hancock, P.A., & Rahimi, M. (1989). Techniques of mental workload assessment. In: J. Wilson (Ed.). *Evaluation of human work*: *practical ergonomics methodology*. (pp.749-782). London: Taylor & Francis.

O'Donnell, R.D., & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.), *Handbook of human performance. Vol. 2. Cognitive processes and performance (*pp. 42.1-42.49). New-York: Wiley.

Parasuraman, R., & Hancock, P.A. (2001). Adaptive control of cental workload. In: P.A. Hancock and P.A. Desmond, (Eds.), *Stress, workload, and fatigue* (1st ed., pp 305-320). Mahwah, NJ: Lawrence Erlbaum Associates.

Reid, G.B., & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In: P.A. Hancock and N. Meshkati (Eds.) *Human mental workload.* (pp. 185-218). New York: North Holland.

Rubio, S., Diaz, E., Martin, J., & Puente, J.M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review*, *53*(1), 61-86.

Szalma, J.L., Warm, J.S., Matthews, G., Dember, W.N., Weiler, E.M., Meier, A., & Eggemeier, F.T. (2004). Effects of sensory modality and task duration on performance, workload, and stress in sustained attention, *Human Factors, 46(2),* 219-233.

Vidulich, M.A. & Tsang, P.S. (1986). Techniques of subjective workload assessment: a comparison of SWAT and the NASA-bipolar methods. *Ergonomics*, *29*, 1385-1398.

Vidulich, M.A. & Wickens, C.D. (1986). Causes of dissociation between subjective workload measures and performance. Caveats for the use of subjective assessments. *Applied Ergonomics*, *17*, 291-296.

Warm, J.S., Dember, W.N., Gluckman, J.P., & Hancock, P.A. (1991). Vigilance and workload. *Proceedings of the Human Factors Society*, *35*, 980-981.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In: R. Parasuraman and M. Mouloua (Eds.), *Automation and human performance.* (pp. 183-200). Mahwah, NJ: Erlbaum.

Wickens, C.D. (1980). The structure of attentional resources. In: R.S. Nickerson (Ed.), *Attention and performance VIII* (pp. 239-257). Erlbaum. Hillsdale, N.J.

Wierwille, W.W., & Cascali., J.G. (1983). A validated rating scale for global mental workload measurement applications, *Proceedings of the Human Factors Society, 27,* 129-133.

Wierwille, W.W., and Eggemeier, F.T. (1993). Recomendations for mental workload measurement in a test and evaluation environment. *Human Factors, 35*(2), 263-281.

Yeh, Y. Y., & Wickens, C. D. (1988). Dissociations of performance and subjective measures of workload. *Human Factors, 30,* 111-120.

Yerkes, R.M., & Dodson, J.D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation, *Journal of Comparative Neurology and Psychology, 18,* 459-482.

Zakay, D. (1998). Attention allocation policy influences prospective timing. *Psychonomic Bulletin & Review*, *5*(1), 114-118.

Zakay, D., & Block, R.A., (2004). Prospective and retrospective duration judgments: An executive-control perspective, *Acta Neurobiolgiae Experimentalis, 64,* 319-328

ACKNOWLEDGMENT

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621, P.A. Hancock, Principal Investigator. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors would also like to express their gratitude to the ninety-two police officers who volunteered to participate in the study for the sole purpose of contributing to a research study. A. Greenwood-Ericksen, J.F. Morgan and J.M. Ross, who helped in data collection and reduction, are also acknowledged.

Table 1

Matrix of performance and workload association and dissociations adapted from Parasuraman, R. & Hancock, P.A. (2001). Adaptive control of mental workload. In: P.A. Hancock and P.A. Desmond (Eds.) Stress, workload and fatigue, (pp.308) Erlbaum: Mahwah, NJ.

Workload		Performance		
	Increase	Association	Performance Insensitivity	Dissociation
	No Change	Workload Insensitivity	Control	Workload Insensitivity
	Decrease	Dissociation	Performance Insensitivity	Association
		Degradation	No change	Improvement

Table 2

Predictions of workload and performance relationships based on the Hancock and Warm extended-U model

Region	Workload	Performance	Relationship
(a)	⇧	⇧	Dissociation
(b)	No change	No change	Control
(c)	⇧	No change	Performance Insensitivity
(d)	⇧	⇩	Association
(e)	Plateau 100%	⇩	Workload Insensitivity
(f)	⇩	⇩	Dissociation

Table 3

Descriptive statistics and confidence intervals on the percentage of hits for each task

Task	Mean*	SD	95% Confidence interval		Mean**	SD	95% Confidence interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
WT	74.7	18.7	70.3	79.2	75.5	19.1	70.6	80.3
FT	65.6	21.6	60.5	70.7	65.7	22.1	60.1	71.3
BT	58.3	20.3	53.5	63.1	58.3	20.9	53.0	63.6
MT	63.0	15.2	59.4	66.6	62.1	15.2	58.2	65.9

*N=71, **N=62 (participants with a complete data set of Performance and Workload data)

Table 4

Descriptive statistics and confidence intervals on global workload scores for each task

Task	Mean*	SD	95% Confidence interval	
			Lower Bound	Upper Bound
WT	43.0	14.9	39.2	46.8
FT	49.2	16.2	45.0	53.3
BT	60.6	15.8	56.6	64.6
MT	57.2	17.0	52.9	61.6

N=67, 68, 68, 69 accordingly, Valid (N)=62

Table 5

NASA-TLX Subscale Descriptives for the Four Task (WT, FT, BT, and MT)

Task	WT		FT		BT		MT	
TLX element	Mean (SD)	'baseline'	Mean (SD)	Relative to 'baseline'	Mean (SD)	Relative to 'baseline'	Mean (SD)	'Relative to 'baseline'
Mental subscale	52 (26)	100	59 (23)	113	68 (21)	131	65 (24)	125
Physical subscale	28 (23)	100	40 (26)	143	62 (24)	221	53 (25)	189
Temporal subscale	38 (26)	100	45 (26)	118	67 (22)	176	62 (25)	163
Performance subscale	63 (24)	100	61 (23)	97	55 (23)	87	52 (22)	83
Effort subscale	48 (23)	100	53 (24)	110	65 (20)	135	62 (21)	129
Frustration subscale	48 (23)	100	38 (23)	79	47 (27)	98	49 (25)	102

N=62. Note than when calculating the Global RTLX the Performance subscale is reversed.

Figure Captions

Figure 1. The extended-U model of stress and performance proposed by Hancock and Warm (1989). See Table 2 for predictions on workload-performance relationships for each region. Note that task demand and dynamic changes in task demands are a major source of stress in this study, and therefore the stress level axis in this model can also be labeled as task demand.

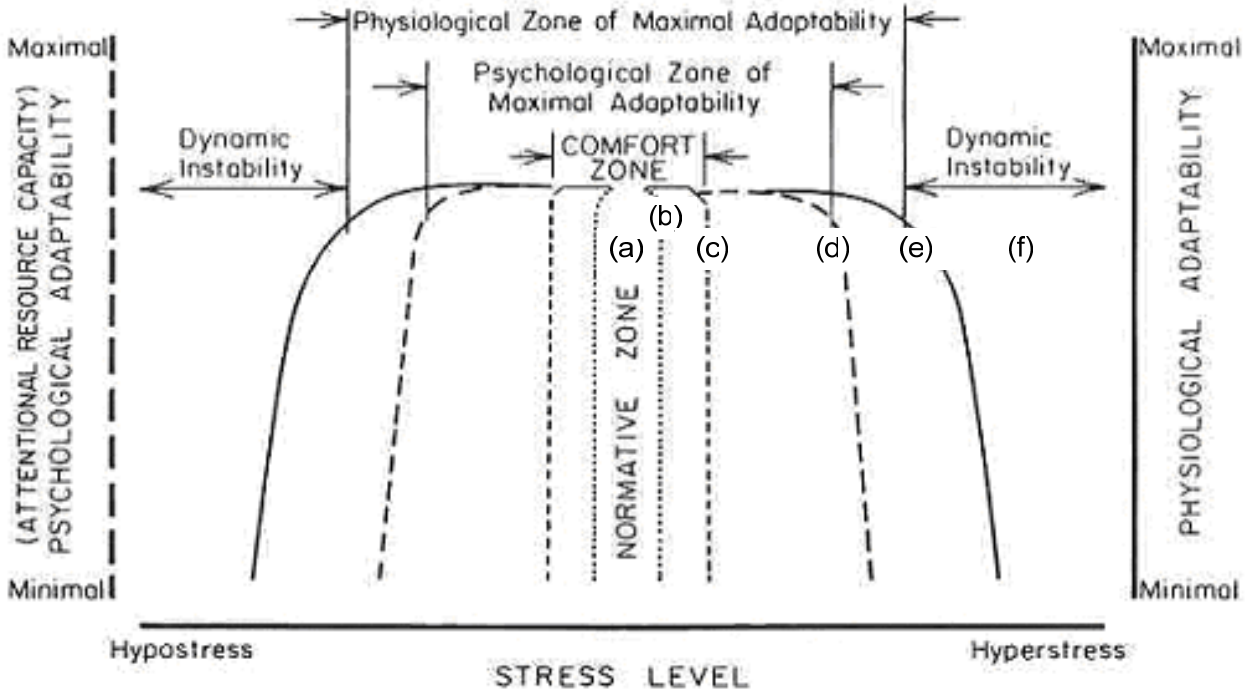


Figure 1

Training for Vigilance: Using Predictive Power to Evaluate Feedback Effectiveness

James L. Szalma and Peter A. Hancock

University of Central Florida, Orlando FL

Joel S. Warm, William N. Dember, and Kelley S. Parsons

University of Cincinnati, Cincinnati, OH

Address correspondence to James L. Szalma, Institute for Simulation and Training, University of Central Florida, PO Box 161390, Orlando, FL 32816-1390; jszalma@mail.ucf.edu

Running Head: Knowledge of Results in Vigilance

Keywords: Vigilance; Sustained Attention; Training; Knowledge of Results; Performance

ABSTRACT

Objective

We examined the effects of Knowledge of Results (KR) on vigilance accuracy, and report the first use of positive and negative predictive power (PPP/NPP) to assess vigilance training effectiveness.

Background

Training individuals to detect infrequent signals amongst a plethora of non-signals is critical to success in many failure-intolerant monitoring technologies. KR has been widely used for vigilance training, but the effect of KR-schedule on accuracy has been neglected. Previous research on training for vigilance has used signal detection metrics or hits and false alarms. In this study diagnosticity measures were applied to augment traditional analytic methods.

Method

We examined the effects of continuous knowledge of results (KR) and a partial-KR regimen versus a no-KR control on decision diagnosticity.

Results

SDT analysis indicated that KR induced conservatism in responding but did not enhance sensitivity. However, KR in both forms equally enhanced PPP while selectively impairing NPP.

Conclusion

There is a tradeoff in the effectiveness of KR in reducing false alarms and misses. Together, SDT and PPP/NPP measures provide a more complete portrait of performance effects.

Application

PPP and NPP together provide another assessment technique for vigilance performance, and as an additional diagnostic tool these measures are potentially useful to the human factors community.

INTRODUCTION

Vigilance refers to the ability of observers to detect infrequent signals over prolonged periods on watch (Davies & Parasuraman, 1982). The importance of vigilance has vaulted to the forefront of current social concerns regarding detection of terrorist activities (Hancock & Hart, 2002). For example, the screening devices deployed by the Transportation Security Administration (TSA) to screen 100% of checked baggage represents a major vigilance requirement for operators (Harris, 2002). Sustained attention is also central to the majority of advanced human-machine systems, including; air-traffic control, cockpit monitoring, industrial quality control, nuclear power generation, medical monitoring, cytological screening (Warm, 1993). One perennial concern remains how to train individuals for vigilance and how to evaluate performance. An effective training procedure is to provide information about an operator's responses via knowledge of results (KR). Research has indicated that monitors trained with KR perform better after the training aid has been withdrawn than controls who received no KR training (see Davies & Parasuraman, 1982; Warm & Jerison, 1984). However, the mechanisms by which KR exerts its effects, and how best to employ KR for training, remains unspecified.

In vigilance, performance accuracy is assessed predominantly by evaluating the proportion of correct detections and false alarms. These scores are used to compute signal detection theory (SDT) measures of sensitivity (e.g., d′ or A′; Macmillan & Creelman, 2005; See, Howe, Warm, & Dember, 1995) and response bias (e.g., β; Davies & Parasuraman, 1982; and see See, Warm, Dember, & Howe, 1997). Sensitivity assesses operators' ability to discriminate signal from non-signal. In vigilance, response bias, which tends to become increasingly conservative over time, is thought to reflect increasing awareness on behalf of operators that the signals for which they are searching are relatively rare (Craig, 1978). Indeed, KR has been shown to enhance perceptual sensitivity while increasing β (e.g., see Szalma, Miller, Hitchcock, Warm, & Dember, 1998). However, while these established measures assess critical aspects of performance, they do not directly capture the diagnostic accuracy of response, i.e., the proportion of 'yes' or 'no' responses by individuals that are correct. SDT measures reflect diagnosticity indirectly, since one might infer greater diagnostic accuracy from higher d' scores. It is possible, however, to achieve high sensitivity while making poorer diagnostic decisions

(Parasuraman, Hancock, & Olofinboba, 1997). This is particularly evident when the *a-priori* probability of a signal (base rate) is very low, as in most real-world situations. Indeed, Parasuraman, Hancock, and Olofinboba (1997) demonstrated that under these conditions the accuracy of system response can also be low even when sensitivity is exceptionally high. An operator with a high sensitivity can generate unacceptably large numbers of false alarms and reduce their diagnostic power, which inevitably results in reduced overall system effectiveness.

It is therefore critical in operational settings that observers learn to be highly diagnostic as well as sensitive to perceptual changes in the displays which they are monitoring. Specifically, it is vital that when an operator indicates a signal is present then the signal actually is present. Conversely, when an operator indicates that a signal is not present, it is vital that the signal is really not there. These aspects of performance are captured by the decision theory measures of positive predictive power (PPP), which is the proportion of 'yes' responses that are actually correct, and negative predictive power (NPP), which is the proportion of 'no' responses that are actually correct. The computational formula for PPP is H/(H+FA), where H= number of signals detected (hits) and FA=number of false alarms. The comparable computational formula for NPP is CR/(CR+M), where CR=number of correct rejections and M=number of signals missed. These indices are frequently employed in evaluating decision-making in medicine (e.g., see Linton, 1996), but have rarely been employed in the human factors literature (for an exception, see Getty, Swets, Pickett, & Gonthier, 1995). A perfectly accurate observer would yield a PPP of 1.0; a score of zero would indicate no correct detections and no diagnosticity. Similarly, an observer who correctly rejected all non-signals and committed no misses would achieve a NPP score of 1.0, while a score of zero indicates that no correct rejections were made.

Given the importance of training personnel for monitoring and the relative costs of false alarms and misses, it is vital that any training regimen maximize both sensitivity and diagnostic power. Hence, we applied PPP and NPP to the examination of the effectiveness of KR for vigilance training, and we compared these measures to non-parametric measures of SDT. We do not propose PPP/NPP measures replace SDT, which is one of the most powerful and useful quantitative theories in behavioral science. Rather, we show

how inclusion of these 'new' measures complements SDT indices and provides an alternative vista into detection capability.

Knowledge of results has been shown to be an effective training aid for a variety of tasks beyond vigilance, including those entailing motor skills (e.g., Salmoni, Schmidt, & Walter, 1984). However, the schedule of KR presentation influences the effectiveness of such training. Specifically, partial-KR, i.e., KR provided only during portions of training, has been shown to be more effective than continuous-KR in training motor skills (Salmoni, Schmidt, & Walter, 1984). If these findings extend to vigilance, performance during training should be superior for monitors receiving continuous-KR relative to those who receive partial-KR, but the pattern should reverse during the transfer phase. Such a result would suggest that partial-KR would be the more effective training mode for vigilance in operational settings.

Most vigilance studies employing KR for training have used a continuous-KR schedule, but a limited number of studies have employed partial-KR (i.e., McCormack, Binding, & McElheran, 1963; Warm, Hagner, & Meyer, 1971). In the latter work, Warm, Hagner, and Meyer (1971) found that observers provided with KR only 50% of the time during training showed a smaller vigilance decrement in a subsequent test period without KR than observers who had received continuous-KR during training. Warm et al. attributed this partial-KR advantage to the motivational effects of feedback. However, they employed a speed measure, *reaction time to correct detections*, to assess performance in a simple vigilance task -- the onset of a small jewel light, and thus no study has examined the effect of partial-KR on vigilance in which the *accuracy* of operator responses was the critical measure of interest. While speed is of importance in detection tasks, it is arguable that accuracy is a more important real-world imperative. Indeed, prior research has shown that accuracy and response time may not always reflect the same perceptual processes (Santee & Egeth, 1982). Thus, effects of partial-KR on response accuracy in vigilance have yet to be evaluated and this is another goal, unique to the present study.

METHOD

Sixteen observers (8 men and 8 women) were assigned at random to each of three KR groups

(continuous-KR, partial-KR, and a no-KR control). Observers ranged in age from 18 to 40 (mean= 20.5), and served to fulfill a University course requirement. All observers had normal or corrected-to-normal vision and were reportedly free of hearing impairments. They took part in a 45-min session divided into a training phase of three continuous 4-min periods of watch followed by a test phase of five continuous 4-min periods. A ten-minute rest interval separated training and testing phases.

The display consisted of a 1.4 cm diameter green disk flanked on each side by a 1.0 cm vertical green line appearing against a gray background. The lines were connected to the disk by a 1.0 cm horizontal green line that extended from the midpoint of each vertical line through the horizontal diameter of the disk. Neutral events, requiring no overt response, were cases in which the two vertical lines were equidistant from the disk. Critical signals were cases in which one vertical line was 0.4 cm farther from the disk than the other line. Observers were instructed to respond by pressing the spacebar on a computer keyboard whenever a critical signal appeared on the screen. Responses occurring within 1.3 sec of critical signal onset were recorded by the computer as correct detections. All other responses were recorded as false alarms. The screen was mounted on a table at eye level approximately 54 cm from the seated observer. This task was selected in part because of existing comparable results readily available to the authors (e.g., Szalma et al., 1998).

In both experimental phases, the disk/line display was presented in the center of the screen for 200 msec once every 2 sec (event rate = 30/min). Critical signals occurred twice/min (signal probability = 0.066) at random intervals within each period on watch. Half of the signals in each period were cases in which the left vertical line was farther from the disk, and the other half were cases in which the right vertical line was farther from the disk. The order in which the two types of critical signal were presented (left vs. right) was randomly distributed over each watch period. During training, KR was provided by the computer using a pre-recorded female voice to announce correct detections ("Correct"), missed signals ("Miss"), and errors of commission ("False Alarm"). In the No-KR condition, the voice announced "saved" after each response. In the partial-KR group, participants were informed that KR would be provided by means of the voice which announced “feedback on” at the beginning of periods 1 and 3, and “feedback off” at the beginning of period

2. Observers in the continuous-KR condition were informed that they would receive feedback throughout the training session. Note that the difference between the continuous- and partial-KR conditions was in the number of periods during which KR was provided rather than the information imparted by the feedback.

RESULTS

Correct Detections

Training Phase. The mean percentages of correct detections during both training and transfer are reported in Table 1. For *training*, analysis of variance (ANOVA) confirmed a significant KR effect, $F(2,45)=8.18$, $p<.01$, $\omega^2 =.09$. The effect for the other factors lacked statistical significance. In this and all subsequent analyses proportions were converted to arcsines to normalize the data. The means and standard errors reported are based on the untransformed percentages. Further, in all analyses, Box's correction was employed to adjust for violations of sphericity. Tukey HSD comparisons indicated that the detection scores of observers in the no-KR condition were significantly higher than those of observers who received either form of KR, but that detection scores in the two KR conditions did not differ significantly from one another.

Transfer Phase. During the *transfer* phase observers in the no-KR group attained higher detection scores than those in the KR groups. An ANOVA indicated a significant main effect for KR, $F(2,45)= 12.00$, $p<.001$, $\omega^2 =.08$, for periods, $F(3,152)= 9.02$, $p<.001$, $\omega^2 =.12$, and for the interaction, $F(7,152)=2.04$, $p<.05$, $\omega^2 =.05$. Simple effects tests of periods within each KR condition indicated a significant decline over time for observers in the continuous-KR group, $F(4,53)=5.89$, $p<.01$, $\omega^2 =.20$, and the no-KR group, $F(3,44)=6.18$, $p<.001$, $\omega^2 =.20$. Detection scores for observers in the partial-KR group did not significantly change over time.

<Insert Table 1>

False Alarms

Training Phase. The mean percentages of false alarms during the training and transfer sessions are reported in Table 1. ANOVA revealed significant effects for KR, $F(2,45)= 9.37$, $p<.001$, $\omega^2 =.10$, and for periods, $F(2,88)= 6.75$, $p<.01$, $\omega^2 =.07$. The interaction between these factors was not statistically significant.

Tukey HSD comparisons indicated that participants in the no-KR condition committed significantly more false alarms than those in either KR condition, but scores in the two KR groups did not differ significantly from one another. Across the KR conditions false alarms declined over periods.

Transfer Phase. During *transfer*, observers in the no-KR control group committed substantially more false alarms than those who received KR training. In fact, 41% of all the false alarm scores were zero for the KR groups across the five periods of watch. An ANOVA indicated a significant effect for KR, F (2,45)= 13.16, p<.001,ω^2 =.09, and a significant interaction, F (7,166)= 2.24, p<.05, ω^2 =.04. The period effect was not significant. Simple effects tests of periods within each KR condition indicated that false alarms committed by observers in the no-KR condition declined significantly over the watch period, F(2, 33)=3.52, p<.05, ω^2 =.11. False alarm scores for observers in the two KR conditions did not change significantly over time.

Signal Detection Analysis

The substantial number of zero false alarm rates in the KR groups rendered application of parametric SDT analysis inadvisable due to the problems associated with computation of such indices under these conditions (see Davies & Parasuraman, 1982; Macmillan & Creelman, 2005). To circumvent these concerns Craig (1979) suggested that non-parametric sensitivity indices (specifically, A′) should be used when parametric measures are inappropriate. Work by See et al. (1997) established that among the available, non-parametric indices of response bias, β''_D was the most effective for vigilance. The correct detection and false alarm scores for each participant were therefore used to compute A′ and β''_D.

Perceptual Sensitivity (A′)

Training Phase. An ANOVA of the A′ scores during *training* indicated that sensitivity did not significantly differ across the three KR conditions, and that A′ scores for observers remained stable with time on watch. The interaction between these factors also lacked significance.

Transfer Phase. Sensitivity scores during *transfer* for the three KR conditions are plotted as a function of periods of watch in Figure 1a. An ANOVA indicated a significant effect for KR, F (2,45)= 6.82, p<.01, ω^2 =.05, and a significant effect for period, F (4,104)= 4.20, p<.05, ω^2 =.05. Sensitivity of observers declined over time in all conditions, confirming the classic "vigilance decrement" (See et al.,1995). The interaction was not statistically significant. Tukey HSD tests indicated that observers in the continuous-KR group (M=.85; d=-.78)

were significantly less sensitive than those in the no-KR condition (M=.91). A′ scores for observers in the partial-KR condition (M=.88; d=-.53) did not differ significantly from those in the other two groups.

<Insert Figure 1 about here>

Response Bias (β_D'')

Training Phase. An ANOVA on β''_D scores during *training* indicated a significant effect for KR, F (2,45)= 13.93, p<.001 ω^2 =.15. However, there were no significant differences in β''_D scores across periods, and the interaction was also not statistically significant. Tukey HSD tests revealed that observers in the continuous-KR (M=.94; d=.84) and partial-KR (M=.95; d=.87) groups did not significantly differ from one another, but both groups were significantly more conservative than those in the no-KR condition (M=.54).

Transfer Phase. Mean response bias scores for observers in the three KR conditions during *transfer* are plotted as a function of periods of watch in Figure 1b. Observers who received either form of KR achieved a conservative level of responding early in the watch and remained at that level, while observers who did not receive KR were more lenient in responding and became more conservative over time. ANOVA revealed a significant effect for KR, F (2,45) = 13.21, p<.001, ω^2 =.09, for period, F (4,77) = 7.92, p<.01, ω^2 =.10, and for the interaction, F (8,77)= 6.85, p<.001, ω^2 =.16. Tests for the simple effects of period within each KR condition confirmed that observers in the no-KR group became more conservative over time, F (4,77)= 21.49, p<.01, ω^2 =.51, but that β''_D scores of observers in the two KR groups did not change significantly over time.

Diagnosticity Measures

The PPP and NPP were computed for each participant in each period of watch during both training and transfer sessions. It is important to remember that in vigilance the number of non-signals far exceeds the number of signals, so that opportunities for errors of omission are much rarer than opportunities for false alarms. Thus, in cases where there is a low signal probability and many correct rejections (i.e., the observer adopts a conservative criterion), NPP values would be expected to be artificially high (above 90%). Therefore, any differences in NPP among KR conditions will likely be driven by differences in errors of omission. In contrast, PPP is influenced by frequencies of correct detections and false alarms, and these responses are relatively few in number compared with the total number of events during a vigil.

Positive Predictive Power (PPP)

Training Phase. An ANOVA of the PPP scores revealed a significant period effect, $F(2,83)= 5.50$, $p<.01$, $\omega^2 =.06$, with scores increasing over time (M's for periods 1-3= .62, .68, and .71, respectively). No significant differences among KR groups were observed, and the interaction also lacked statistical significance. *Transfer Phase*. Mean PPP scores for the three groups are displayed as a function of period on watch in Figure 2a. ANOVA revealed a significant effect for KR, $F(2,45)= 6.92$, $p<.01$, $\omega^2 = .05$. No significant differences were observed for period or the KR by period interaction. Tukey HSD tests revealed that the two KR groups did not significantly differ from one another (continuous-KR M=.83; d=.96; partial-KR M=.81; d=.88), but the PPP scores of both these groups exceeded those of the no-KR group (M=.63).

<Insert Figure 2 about here>

Negative Predictive Power (NPP)

Training Phase. An ANOVA revealed a significant effect for KR group, $F(2,45)= 11.69$, $p<.01$, $\omega^2 =.13$. No significant effects were observed for periods or the interaction between these factors. Tukey HSD tests revealed that the no-KR group (M=.98) achieved NPP scores that were significantly higher than those of the continuous-KR (M=.97; d=.72) and partial-KR groups (M=.96; d=.99). The NPP scores of the two KR conditions did not differ significantly from one another. Although these absolute mean differences are small, the effect sizes are large as a result of low within-group variability.

Transfer Phase. Mean NPP scores for the three groups are displayed as a function of period on watch in Figure 2b. Observers in the no-KR condition achieved higher NPP scores during transfer than those in the two KR conditions. ANOVA indicated significant effects for KR, $F(2,45)= 12.20$, $p<.001$, $\omega^2 = .09$, periods, $F(3,141)= 10.93$, $p<.001$, $\omega^2 = .14$, and the interaction, $F(6, 141)= 2.82$, $p<.05$, $\omega^2 = .06$. Simple effects tests of periods within each KR condition indicated that NPP scores significantly declined over the watch period in the no-KR condition, $F(3,141)= 11.07$, $p<.01$, $\omega^2 =.14$, and the continuous-KR condition, $F(3,141)= 4.30$, $p<.01$, $\omega^2=.05$. NPP scores for participants in the partial-KR condition did not change significantly over time.

DISCUSSION

Signal Detection

In this experiment we found no evidence that KR enhanced perceptual sensitivity. This was surprising, given that KR typically enhances signal detection in vigilance (Warm & Jerison, 1984). However KR did increase conservatism in responding relative to a no-KR condition, a finding that has been observed previously in vigilance experiments (See et al., 1997). The increase in conservatism over time on watch in the no-KR condition is also consistent with prior research, and may be due to the growing awareness over the vigil that critical signals are relatively rare (Craig, 1978). The provision of KR likely produces an immediate and explicit awareness of the rarity of signals and so causes observers to set a high criterion very early in the vigil.

There was a tradeoff between PPP and NPP such that the enhanced diagnosticity provided by KR for 'yes' responses (PPP) is achieved at the cost of lower diagnosticity for 'no' responses (NPP). Consideration of both the SDT and diagnosticity metrics indicates that KR increases conservatism and improves the reliability of 'yes' responses, but that this occurs at the expense of lower accuracy in 'no' decisions and lower perceptual sensitivity. That is, KR facilitated learning to avoid false alarms, reflected in enhanced PPP and increased conservatism, but did not facilitate learning to avoid misses, manifested in lower levels of NPP and sensitivity.

The differential efficacy of KR for reducing false alarms versus misses may have resulted in part from

the perceptual discrimination required. As in most vigilance experiments, stimuli were presented on the screen only for a brief duration. These brief exposure times, coupled with the spatial discrimination required, induced gamma motion in that the two vertical lines appeared to 'move' from the center dot when they appeared on the screen (see Figure 1). This apparent motion may have induced leniency in responding which was reduced by the feedback. However, this task feature may have distracted observers in the KR conditions from learning the distances to be discriminated. Apparent motion did not affect vigilance per se, since the no-KR control results replicate those typically observed. Rather, the apparent motion may have interacted with the informational value of each form of KR which subsequently interfered with the learning process in regard to sensitivity.

Evidence that different forms of KR are not equivalent in their informational value or in their performance effects comes from the work of Dittmar, Warm, and Dember (1985). They provided observers with feedback regarding correct detections, false alarms, or misses. While hit and false alarm KR enhanced detection accuracy, miss KR *failed* to enhance observers' detection ability. These effects occurred despite the fact that hit and false alarm KR should, in principle, provide observers with the same informational feedback.

The relative effectiveness of KR for reducing false alarms versus misses may also depend on whether the feedback occurs after overt responding. In most vigilance experiments overt responses were required only when the observer believed a signal was present (for one exception see Parasuraman & Davies, 1976). Non-signal events therefore require the 'response of not responding,' and no KR is provided for correct rejections. It may be that the KR typically used in vigilance is effective only when observers make an overt response and thereby link the feedback to a specific self-generated event. KR that cannot be so associated with a discrete action (miss KR) may therefore be less effective. Indeed, Dittmar et al. (1985) argued that miss KR does not adequately provide sufficient information to be perceptually useful. Requiring overt responses for non-signals may therefore enhance the efficacy of KR especially for improving NPP and sensitivity. Further exploration of this potentiality is required to establish what elements of KR produce these differential performance effects.

Criterion Setting or Diagnosticity?

The results of the SDT analysis might tempt one to conclude that the effects of KR on PPP and NPP are driven solely by criterion setting. However, reduction in responding *per se* does not need to result in a change in a particular direction of either PPP or NPP. This can be seen by the example in Table 2, which shows hit and false alarm rates that produce common levels of response bias but different values of PPP and NPP, and a comparable example analysis in which A′ is held constant. The diagnosticity measures are determined by the relative frequencies of correct and incorrect decisions, rather than the absolute level of 'yes' responses. Therefore, both the operator's perceptual ability and level of responding can influence diagnosticity.

<Insert Table 2 about here>

KR Schedule

One goal for this study was to investigate the possibility that observers who receive partial-KR training perform better during a test vigil than those who experience continuous-KR training. This hypothesis was not supported, an outcome that differs from that obtained by Warm, Hagner, and Meyer (1971). They reported that partial-KR training facilitated performance during a test vigil. However, Warm and his colleagues evaluated *reaction time* while this study evaluated *response accuracy*. These aspects of performance are thought to reflect different perceptual processes (Santee & Egeth, 1982). Reaction time to easily detectable stimuli, such as those employed by Warm and his colleagues, likely reflects resource-limited processing. This in turn may be influenced by motivational processes. By contrast, the present detection task required a much more difficult discrimination and is much more likely to be data-limited and thus not influenced by increased effort (c.f., Norman & Bobrow, 1975). Partial-KR may be more effective than continuous-KR when performance is in the resource-limited range as participants are able to modulate their effort. They may reduce their effort during the KR portions of the experiment and increase effort during the portions in which no feedback is provided. In the data-limited range, effort modulation will not influence performance, rendering the two forms of KR equally efficacious. This interpretation is consistent with an effort-regulation model in which effort level is modulated according to the level of demand (see Hockey,

1997). Modulation of effort may be effective when the task is resource-limited, but in the data-limited range such regulation is less flexible and therefore less beneficial.

SDT and PPP/NPP

The effects of KR and time on task in this study were contingent upon the dependent measures employed. A′ showed a decline over time and a negative effect of KR during transfer, and β''_D showed a KR effect during training and an interaction during transfer. PPP showed no time effects during either session, but did show a facilitative effect for KR during transfer. The NPP analysis indicated a debilitative effect of KR during training and an interaction between KR and time during transfer. Thus, the two sets of measures did not produce the same patterns of outcome. The question is not which set of metrics is better, but rather how the metrics may be used together to provide a more complete performance portraiture. Sensitivity provides information regarding stimulus discriminability, while response bias reflects the decision criterion. These measures correspond more directly to psychological processes although they do not imply specific mechanisms. Indeed, Swets (1977) argued that there is more to vigilance than discrimination and criterion setting, and that SDT cannot provide a total explanation for its performance effects. Further, sensitivity measures often reflect more than pure sensory processing ability (Pastore, Crawley, Berens, & Skelly, 2003). So, while SDT indices may tend to reflect psychological processes they do not, for this reason alone, supersede other useful measures. The PPP/NPP approach provides a direct assessment of the accuracy of yes/no responses given particular levels of sensitivity and bias, and thereby indicates how often the individual's decision will be correct. Given that poor diagnostic decisions can occur when signal base rates are low (Parasuraman et al., 1997), measures that directly reflect diagnostic power can be especially useful in evaluating real-world detection systems.

Practical Applications

The Transportation Security Administration (TSA) has placed over 1 billion dollars of detection equipment into operation at over 400 airports within the United States (Airports Staff, 2001). The efficiency of such systems depends crucially upon the capability of operators who use them. To enhance operator

efficiency technologies such as the Threat Image Projection System (TIP; Neiderman & Fobes, 1997) overlay computer-generated images on to inspected items to provide feedback for performance improvement. Here we demonstrate that the role of feedback in vigilance can be interpreted differently, contingent on the nature of the dependent measure employed. The relative importance of diagnosticity for signal and non-signal events is therefore critical in evaluating the effectiveness of KR manipulations as training strategies for monitoring tasks. Our results show that KR effectiveness depends in part on whether emphasis in training is on false-alarm reduction or on the reduction of missed signals. For the latter, it might be advisable to require operators to make overt 'no' responses during training to attenuate the decline in the accuracy.

Aviation security issues are only one example among many domains in which training for vigilance has practical utility. Another obvious realm of concern is detection and decision making in medicine. While PPP and NPP are relatively new to human factors professionals, these measures are often used by health professionals to whom the human factors community is now reaching out (Bogner, 1994). Therefore, we must assure those professionals with whom we wish to interact that we can embrace and integrate their measurement approaches with those which we are already comfortable and familiar. This study is the first direct comparison between PPP/NPP with the classic measures of signal detection in sustained attention. Our results have shown that PPP/NPP provide another window on performance which is especially useful when false alarm rate is naturally low. We hope that others will see the advantage of these differing methods and, given the simple calculations needed for their derivation, include PPP and NPP in their own efforts to understand operators' capabilities to detect environmental signals.

ACKNOWLEDGEMENT

Completion of this work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. P.A. Hancock, *Principal Investigator.* The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant. We are also grateful to Eric Warm, M.D. for helpful discussions regarding the application of diagnosticity measures in medicine.

REFERENCES

Airports Staff. (2001, November 15). Airports: S.1447, Aviation and Transportation Security Act. *Aviation Week and Space Technology*. Retrieved January 29, 2004 from http://www.aviationnow.com

Bogner, M.S. (1994).(Ed.). *Human error in medicine*. Hillsdale, NJ: Erlbaum.

Craig, A. (1978). Is the vigilance decrement simply a response toward probability matching? *Human Factors*, *20,* 441-446.

Craig, A. (1979). Nonparametric measures of sensory efficiency for sustained monitoring tasks. *Human Factors, 21*, 69-78.

Davies, D.R., & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Dittmar, M.L., Warm, J.S., & Dember, W.N. (1985). Effects of knowledge of results on performance in successive and simultaneous vigilance tasks: A signal detection analysis. In: R.E. Eberts and C.G. Eberts

(Eds.), *Trends in ergonomics/human factors II.* (pp. 195-202). Amsterdam: Elsevier.

Getty, D.J., Swets, J.A., Pickett, R.M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1,* 19-33.

Hancock, P.A., & Hart, S.G. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in Design, 10,* 6-16.

Harris, D. (2002). How to really improve airport security. *Ergonomics in Design*, *10*, 17-22.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*, 73-93.

Linton, C.S. (1996). General internal medicine. In: U.B.S. Prakash (Ed.), *Mayo Internal Medicine Board Review 1996-1997.* (pp. 333-352). Rochester, MN: Mayo Foundation for Medical Education and Research.

Macmillan, N.A., & Creelman, C.D. (2005). *Detection theory: A user's guide*. (2nd Edition). Mahwah, NJ: Erlbaum.

McCormack, P.D., Binding, F.R.S., & McElerhan, W.G. (1963). Effects on reaction time of partial knowledge of results of performance. *Perceptual & Motor Skills, 17*, 279-281.

Neiderman, E.C. & Fobes, J.L. (1997). Operational test and evaluation of human factors interventions for airport security screeners. *Proceedings of the Human Factors and Ergonomics Society*, *41*, 1094-1097. Santa Monica, CA: Human Factors and Ergonomics Society.

Norman, D.A., & Bobrow, D.G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

Parasuraman, R., & Davies, D.R. (1976). Decision theory analysis of response latencies in vigilance. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 578-590.

Parasuraman, R., Hancock, P.A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision warning systems. *Ergonomics, 39*, 390-399.

Pastore, R.E., Crawley, E.J., Berens, M.S., & Skelly, M.A. (2003). "Nonparametric" A' and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review, 10*, 556-569.

Salmoni, A.W., Schmidt, R.A., & Walter, C.B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin, 95*, 355-386.

Santee, J.L., & Egeth, H.E. (1982). Do reaction time and accuracy measure the same aspects of letter recognition? *Journal of Experimental Psychology: Human Perception and Performance, 8*, 489-501.

See, J.E., Howe, S.R., Warm, J.S., & Dember, W.N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117*, 230-249.

See, J.E., Warm, J.S., Dember, W.N., & Howe, S.R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors, 39*, 14-29.

Swets, J.A. (1977). Signal detection theory applied to vigilance. In: R.R. Mackie (Ed.), *Vigilance: Operational performance and physiological correlates.* (pp. 705-718). New York: Plenum Press.

Szalma, J.L, Miller, L.C., Hitchcock, E.M., Warm, J.S., & Dember, W.N. (1998). Intraclass and interclass transfer of training for vigilance. In: M.W. Scerbo and M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 183-187). Mahwah, NJ: Erlbaum.

Warm, J.S. (1993). Vigilance and target detection. In: B.M. Huey and C.D. Wickens, (Eds.), *Workload transition: Implications for individual and team performance.* (pp.139-170). Washington, DC: National Academy Press.

Warm, J.S., Hagner, G.L., & Meyer, D. (1971). The partial reinforcement effect in a vigilance task. *Perceptual & Motor Skills, 32*, 987-993.

Warm, J.S., & Jerison, H.J. (1984). The psychophysics of vigilance. In: J.S. Warm (Ed.), *Sustained attention in human performance.* (pp. 15-59). Chichester, United Kingdom: Wiley.

Table 1. Mean percentage of correct detections and false alarms during training and transfer for three KR conditions

	Period on Watch						
Correct Detections	1	2	3	4	5	Mean	Cohen's d
Training							
No KR	67 (7)	70 (4)	67 (6)			68 (4)	
Continuous KR	62 (5)	52 (5)	50 (3)			54 (4)	.88
Partial KR	55 (5)	44 (5)	44 (4)			48 (4)	1.25
Mean	61 (3)	56 (3)	54 (3)				
Transfer							
No KR	77 (5)	74 (5)	65 (6)	64 (5)	63 (4)	69 (4)	
Continuous KR	61 (.04)	45 (4)	46 (5)	41 (5)	38 (4)	46 (3)	-1.53
Partial KR	62 (4)	50 (4)	55 (4)	54 (4)	58 (4)	56 (2)	-.98
Mean	67 (3)	56 (3)	55 (3)	53 (3)	53 (2)		
False Alarms							
Training							
No KR	4.4 (.7)	4.6 (.7)	3.5 (.7)			4.1 (.4)	
Continuous KR	2.5 (.4)	1.0 (.2)	1.2 (.2)			1.5 (.4)	.94
Partial KR	2.2 (.4)	1.2 (.2)	1.5 (.3)			1.6 (.4)	.90
Mean	3.0 (.3)	2.3 (.3)	2.0 (.3)				
Transfer							
No KR	4.6 (.8)	4.6 (.7)	3.7 (.7)	3.6 (.7)	3.3 (.6)	4.0 (.6)	
Continuous KR	1.2 (.3)	.5 (.2)	.7 (.2)	.9 (.3)	.7 (.2)	.8 (.1)	-1.58
Partial KR	1.2 (.3)	.9 (.3)	.9 (.2)	.9 (.3)	1.7 (.5)	1.1 (.2)	-1.39
Mean	2.3 (.3)	2.0 (.3)	1.8 (.3)	1.8 (.3)	1.9 (.3)		

Note. Values in parentheses are standard errors. Cohen's d was computed by comparing each experimental group to the No KR control condition.

Table 2. Diagnosticity measures corresponding to hit/false alarm pairs with the same response bias level.

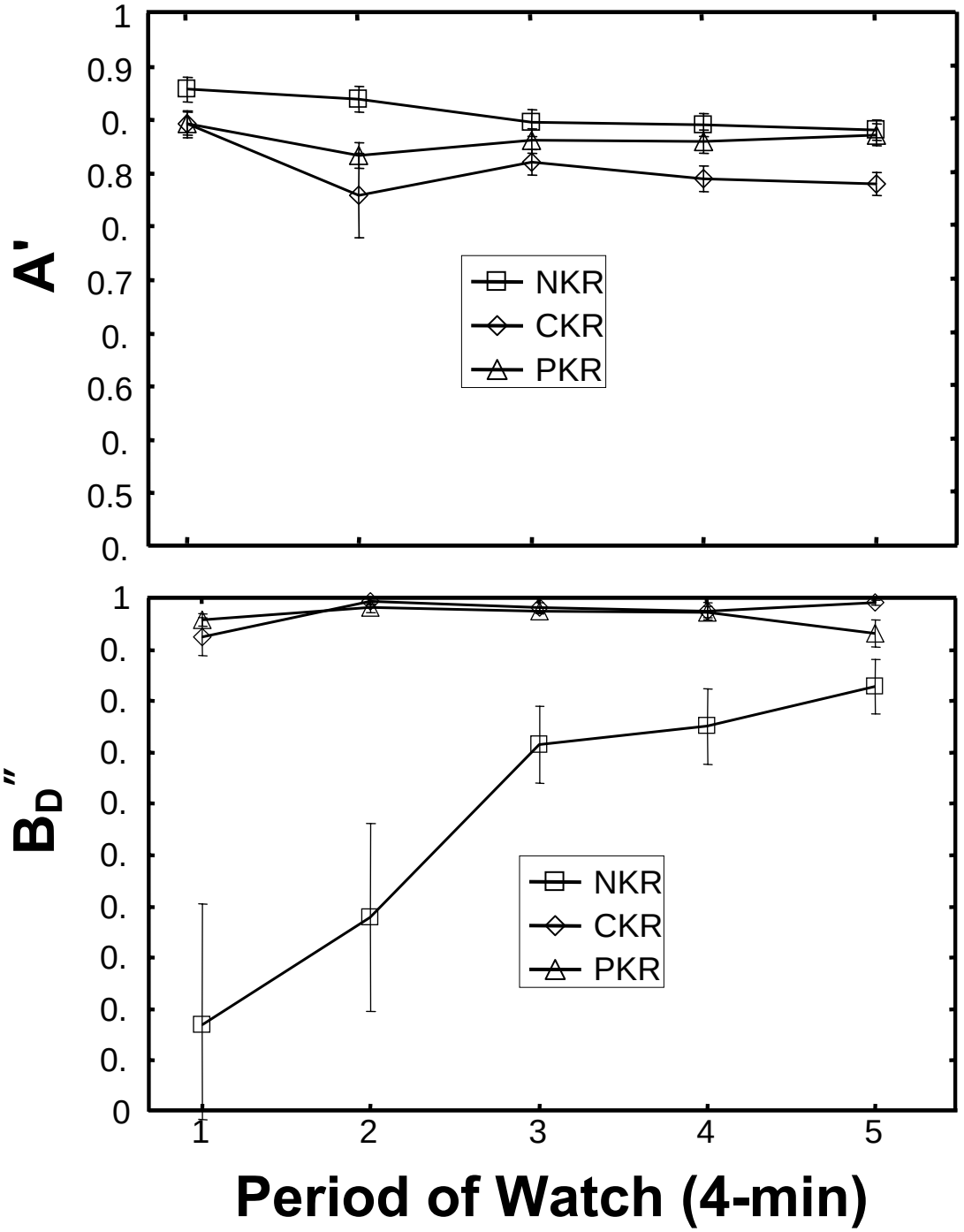
A'	B_D''	p(H)	p(FA)	PPP	NPP
0.50	0.99	0.077	0.053	0.09	0.93
0.50	0.99	0.120	0.031	0.22	0.94
0.51	0.99	0.178	0.017	0.43	0.94
0.52	0.99	0.250	0.009	0.66	0.95
0.59	0.99	0.500	0.004	0.89	0.97
0.70	0.99	0.688	0.002	0.96	0.98
0.80	0.99	0.812	8.93×10^{-4}	0.98	0.99
0.86	0.99	0.875	8.93×10^{-4}	0.99	0.99
0.96	0.99	0.969	1.19×10^{-4}	1.00	1.00
0.71	0.08	0.838	0.143	0.30	0.99
0.71	0.27	0.812	0.116	0.33	0.98
0.71	0.35	0.800	0.107	0.35	0.98
0.71	0.60	0.762	0.071	0.43	0.98
0.71	0.66	0.762	0.061	0.47	0.98
0.71	0.76	0.738	0.046	0.54	0.98
0.71	0.84	0.736	0.031	0.63	0.98
0.71	0.95	0.712	0.010	0.84	0.98
0.71	1.00	0.700	8.93×10^{-6}	1.00	0.98

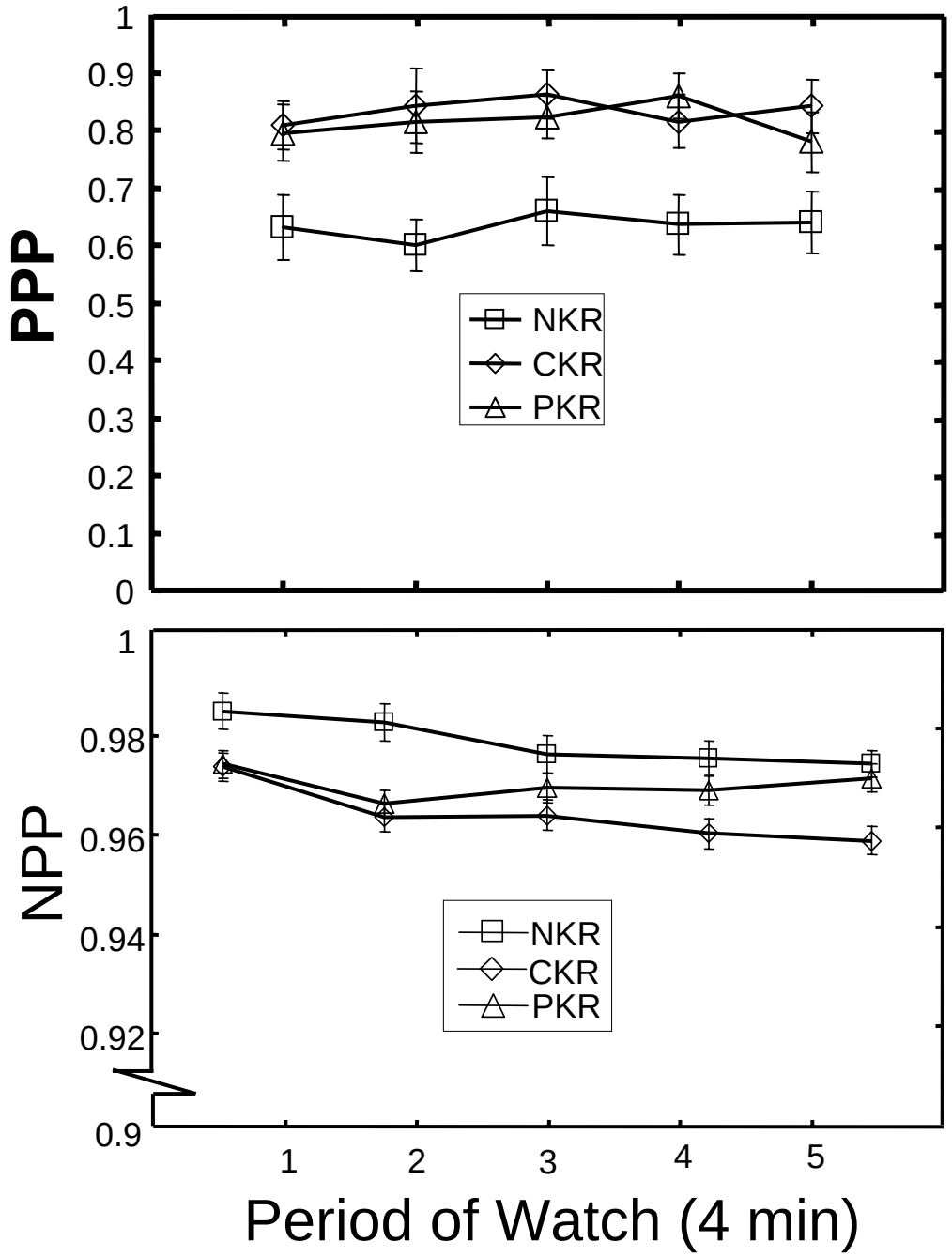
Note. Values in this table assume base rates used in this study (signal rate of .067 and a non-signal rate of .933). PPP=positive predictive power; NPP=negative predictive power.

Figure Captions

Figure 1. (a) Mean sensitivity (A') scores as a function of periods of watch during transfer. (b) Mean response bias (β_D'') scores as a function of periods of watch during transfer. *Note*. CKR = continuous KR; PKR = partial KR; NKR = no KR. Error bars are standard errors.

Figure 2. (a) Mean positive predictive power (PPP) as a function of periods of watch during transfer. (b) Mean negative predictive power (NPP) as a function of periods of watch during transfer. *Note*. CKR = continuous KR; PKR = partial KR; NKR = no KR. Error bars are standard errors.


A'
1
0.9
0.8
0.7
0.6
0.5
NKR
CKR
PKR
B"D
1
0
1
2
3
4
5
Period of Watch (4-min)



Biographies

James L. Szalma is a Research Scientist at the University of Central Florida. He received his Ph.D. in experimental psychology/human factors from the University of Cincinnati in 1999.

Peter A. Hancock is the Provost Distinguished Professor of Psychology at the University of Central Florida, Orlando, FL. He received his Ph.D. in motor performance from the University of Illinois at Urbana-Champaign in 1983.

Joel S. Warm is Professor of Psychology and director of graduate studies in psychology at the University of Cincinnati. He received his Ph.D. in experimental psychology from the University of Alabama in 1966.

William N. Dember is Dean Emeritus and Professor Emeritus at the University of Cincinnati. He received his Ph.D. in experimental psychology from the University of Michigan in 1955.

Kelley S. Parsons is a Graduate Student in the doctoral program in experimental psychology/human factors at the University of Cincinnati. She received an M.A. in experimental psychology/human factors from the University of Cincinnati in 2001.

Are All Teams Created Equal? An Examination of Teamwork Communication across Teams.

Heather A. Priest[2], C. Shawn Burke[2], Joseph Guthrie[2], Eduardo Salas[1,2], and Clint A. Bowers[1]

Department of Psychology[1]

&

Institute for Simulation & Training[2]

University of Central Florida

U.S.A.

DO NOT CITE WITHOUT PERMISSION

Address Correspondence to: Heather A. Priest
Institute for Simulation and Training
University of Central Florida
3100 Technology Pkwy
Orlando, FL 32826-0544
407-882-1347
hpriest@ist.ucf.edu

Keywords: interdependence
team performance
communication
team type

Priest, H. A., Burke, C. S., Guthrie, J., Salas, E. & Bowers, C. A. (under review). Are all teams created equal? An examination of teamwork communication across teams. Submitted to *Human Factors*.

ABSTRACT

"Study Design" This experimental study "Objectives" investigated process differences between teams that varied in level of interdependency, distribution, and shared visual cues and the relationship between team process and team performance. "Summary of Background Data" The majority of previous research on teams has been conducted on one type of team, command-and-control, and results have been generalized to all types of teams. We argue these results and generalizations may be misleading because of the impact variables such as task interdependency has on team coordination. "Methods" One hundred fifty-two participants paired as two-person teams were randomly assigned to one of three team types. Team communication was recorded during the task in order to examine team processes. "Results" Statistical analyses revealed that task interdependency moderated the relationship between teamwork behaviors and team performance ($p=.001$) and between open/closed loop communication sequences and performance ($p=.002$). "Conclusions" These results indicated that there are differences in how teams communicate, what they say, and how often they say things based on their level of interdependency, distribution (expertise and geographical), and the amount of cues they receive in the environment. "Applications" Although further research is needed, potential applications of this research include the design of team training based upon specific team characteristics and structure.

Acknowledgement

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

INTRODUCTION

The past forty years has witnessed much change in the areas of organizational theory, structure, and business practice. Specifically, technological advances, geopolitical stability, and free trade agreements have increased organizational competition within a global economy. In order to remain adaptive and prosper under these circumstances many organizations are witnessing a flattening of traditional, hierarchical structures in favor of teams (Kozlowski & Bell, 2002; Zaccaro, Rittman, Orvis, Marks & Mathieu, 2002). With the increased use of organizational teams has come increased research. This research has taught practitioners and researchers alike much about teams. For example, it is now accepted that teams are characterized by interdependency, adaptive action, and common goals (Salas, Dickinson, Converse, & Tannenbaum, 1992). We also know that both teamwork (i.e., behavioral interactions and attitudinal responses needed to coordinate) and taskwork (i.e., operational) skills are necessary for effective team performance (Morgan, Salas, & Glickman, 1993; Hackman & Morris, 1975; Bass, 1990; Brannick et al., 1993). Furthermore, team research is conducted with the knowledge that teamwork is characterized by a complex set of knowledge, skills, and attitudes (Salas & Cannon-Bowers, 2000; Stevens & Campion, 1994).

As team research has progressed, researchers and practitioners have begun to explore teams more deeply. One of the ways that this has been done is through the creation of team taxonomies. The goal of taxonomies is to define and classify objects or elements so broad statements can be made regarding property classes and class relationships (Fleishman & Quaintance, 1984). The end product of taxonomies is a structure that can provide a common understanding to a diverse body of research.

Team Taxonomies

Over the past several decades, many taxonomic efforts have attempted to classify team phenomenon. Several of these taxonomies have focused on types of teams typically found within work environments (e.g., Argote & McGrath, 1993; Cannon-Bowers et al., 1992; Cohen & Bailey, 1997; Devine, 2002; Devine, Clayton, Philips, Dunford & Melner, 1999; Hackman, 1990; Mohrman, Cohen, & Mohrman, 1995; Sundstrom, McIntyre, Halfhill, & Richards, 2000). Taxonomies focused on team types arose out of an awareness that organizations housed a variety of teams that came in many different shapes and sizes. However, while it is apparent that organizational teams possess different levels of interdependency, the team literature and the corresponding team type taxonomies have traditionally treated interdependence as a dichotomous variable, failing to recognize its continuous nature. One only has to look to operational teams to realize that teams possess different levels of interdependency.

For example, consider a team which is tasked with producing a new computer program. This team possesses a shared goal (i.e., completion of the program) and a team structure in which designers illustrate the program using story boards, story boards are handed off to programmers who create the software, which is then given to usability experts to test. While there is interdependence, it is different than, for example, a team of advertising executive who have a goal to design a new ad campaign for a product. While this team also has a shared goal, member interaction is structured much differently than a team sitting around a conference table brainstorming and pitching ideas.

Task Interdependency

Theoretical literature also offers evidence to support the importance of task interdependency to performance and process. For example, taking a multi-disciplinary perspective and looking outside the team literature, one finds several taxonomies that serve to classify interdependence or at least the work structure of teams based on the tasks teams perform (see Saavedra, Earley, & Van Dyne, 1993; Thompson, 1967; Van de Ven, Delbecq, & Koenig, 1976). One taxonomy that takes an in depth look at interdependency is Saavedra et al.'s (1993) classification. The authors hold that task interdependence stems from the requirements and constraints inherent in the team's task. This typology consists of pooled, sequential, reciprocal, and team interdependency. Pooled interdependence refers to teams in which each member makes a contribution to team output without the need for direct interaction among team members. Team members with pooled interdependence usually have similar roles and the team performance is the sum of individual performances (e.g., a group of word processing specialists; all attend to an inbox and do not pay attention to other team members). Sequential interdependence occurs when one team member must act before another can act. Interdependence is achieved because there is interaction among members, but the work sequence flows only in one direction. Team members with sequential interdependence have different roles and perform different parts of the task in a specific order (e.g., team members in an assembly line task).

Reciprocal interdependence refers to teams who have temporally lagged two-way interactions that require coordination among team members. Team members have different roles, which are externally imposed, and differing expertise and the task order is flexible (e.g., surgical team). Team interdependence refers to team members who jointly diagnose, problem solve, and collaborate to complete a task, while having the freedom to design their

own jobs. This interdependence requires mutual interactions with team discretion to decide the particular course of inputs and outputs among team members, involving an exchange of materials, information, ideas, or other resources (e.g., product development teams). Although not often considered in empirical or practical applications, it has been argued that interdependency will have a significant effect on team performance, especially when examined with other variables, such as team processes (e.g., coordination) and team goals (Faraj, 1998; Perez-Alvarez, 1996; Resick & Bloom, 1997).

Shared Visualization and Distribution

Another variable that has been discussed in recent literature is distribution. Teams can be distributed in a number of ways. Many of today's teams have varying levels of all types of distribution (e.g., geographic, temporal, expertise) due to the increasing complexity of tasks and the emergence of advanced technology. This moderates the effectiveness of teams, how they communicate, and what they communicate. Distributed expertise refers to the situation in which expertise is distributed throughout the team. In this situation, individual team members possess knowledge that is unique; the degree of uniqueness will vary.

When teams are distributed temporally or geographically, they have less of a shared visual work space. Co-located teams, obviously, have high shared visualization. High shared visualization aids in the development of shared situational awareness and shared mental models. Teams with lower levels of shared visualization must compensate, usually through communication, in order to overcome the lack of shared visual cues. Technology can compensate for low shared visualization to a certain degree and the mediating effect of shared visualization has been considered in the design of collaborative technology (e.g., Jasek & Jones, 2001). However, the extent to which physical distribution and distributed

expertise effect teams is still, to a large extent, an unanswered question. Furthermore, the interaction of different types and levels of distribution with other variables (e.g., task interdependency) is virtually unexamined.

While team type taxonomies, task interdependencies, and distribution have been independently examined in the literature, the interaction between team type, task interdependency, distribution, and to the relation to team performance has rarely, if ever, been addressed. The majority of team research has tended to focus rather narrowly on a small subset of team types (e.g., command-and-control) with certain characteristics (e.g., distributed performance arrangements, distributed expertise), and typically dichotomous classifications of team interdependency. This is unacceptable for it ignores the nuances and complexities that exist within teams. A recent search of the empirical literature has revealed that the most used testbeds in team research in the past several decades have been DDD (Dynamic Distributed Decision-Making, Kleinman, Pattipati, Luh, & Serfaty, 1989), TIDE² (Team Interactive Decision Exercise for distributed expertise, Hollenbeck, Sego, Ilgen, Major, Hedlund, & Phillips, 1997), and TANDEM (Tactical Naval Decision Making System; Kozlowski, 1996), all of which are representative of command-and-control teams. These results are then disseminated and applied to collectives of all types, so long as they fit the general definition of teams: two or more individuals interacting interdependently and adaptively towards a common goal (Hollenbeck, Ilgen, Sego, Hedlund, Major, & Phillips, 1995).

Although the current authors have participated in a number of such studies and their application, we now fear that this practice may be misleading in certain circumstances. We argue that task interdependence is an important aspect of team performance because it affects

the level of cooperation of the team, the team's ability to prevent a loss in productivity, and the nature of interpersonal interactions among team members (Saavedra et al., 1993). As an example, we point to Sundstrom, et al. (2000) who put forth a widely adopted team type taxonomy, identifying six major team types and corresponding exemplars. Sundstrom differentiates according to 2 primary characteristics: integration and differentiation. These types are production, service, executive/management, project, action/performing, and advisory/parallel teams. For a description of each team type, see Table 1. However, even within these team types, there is a great deal of variation (See Table 2).

More recently, Devine (2002) integrates and extends several team taxonomies. Devine's review examines teams at a more detailed level, differentiating teams according to seven contextual attribute dimensions: 1) fundamental work cycle, 2) physical ability requirements, 3) temporal duration, 4) task structure, 5) active resistance, 6) hardware dependence, and 7) health risk. For example, while Sundstrom would classify medical, firefighter, and military teams as action teams. Devine distinguishes them as separate team types with differing characteristics. See Table 3. Medical teams are in medical work groups and have functions such as treat and heal. They also have high task structure. Firefighter teams are rescue work groups who have functions such as protect and rescue. They have low task structure. In addition, military teams are military work groups who neutralize and protect, with low to moderate task structure. However, all these teams are action teams under Sundstrom et al.'s taxonomy. Devine's work builds upon prior team taxonomies and illustrates the importance of clearly delineating team types, which have different tasks, functions, and task structure. We believe that this leads to different levels of interdependency and different performance requirements. If Sundstrom's classification is examined alone, the

level of clarity identified by Devine (i.e., task, function, structure) is lost. Therefore, the sub components of 'action/performance' teams are lost.

While there is no evidence that teams with varying interdependence or distribution perform differently or utilize divergent team processes to perform effectively, there is no empirical evidence that they do not. We believe that these different "team types" may need to tap different team processes to perform effectively. For example, distributed teams may need to communicate more in order to perform effectively because they are limited in the amount of nonverbal cues they have access to in the environment. We believe that it may no longer be effective to assume that results from command-and-control teams can be generalized to teams of all types.

In an effort to begin to investigate how communication and teamwork, as indexed via communication context, might differentially relate to team performance across teams with different structural characteristics we created a three level testbed representing three distinct team configurations. This configuration was based in part on established team type taxonomies. We first developed a practical, integrative framework that includes team type, task characteristics, and interdependency as a continuous variable (see Figure 1) to guide our testbed development and experimental plan. The most important aspect of this framework for current purposes is both the continuous nature of interdependency and the interaction of team type (see Table 1; Sundstrom et al., 2000), task characteristics (e.g., co-located environment, expertise), and interdependency (see Table 4; Saavedra et al., 1993). This framework served as a basis for the experimental work described below, meeting three specific goals. First, it served as a practical mechanism by which to systematically categorize what we really know about teams after forty years. Our hypothesis was that most of the research has been

conducted on a limited sample of teams with similar levels of interdependency. Second, this framework guided the development of a three level testbed that allowed us to target specific team types according to our framework. Third, it served to facilitate the identification of future research needs in a systematic manner. As mentioned earlier, using the taxonomy, we created 3 different team configurations that correspond to the variables (i.e., team type, task characteristics, interdependency) contained within the matrix.

Communication and Team Performance

Communication Context. Beyond team type taxonomies, the focus of a majority of traditional team research has revolved around performance and the measurable tangibles that teams deliver. Within recent years, however, there has been a move to understand the 'black box' known as team process. This has led to an emphasis on communication. We believe that the importance of communication can not be emphasized enough. Communication often serves as a measurable proxy for the many opaque, "immeasurable" processes that teams need to coordinate and perform effectively. The question remains, however, how do you best measure communication to get at team process? Typically this has been done through analyzing team communication in terms of frequency or quality and relating this to team performance. Though insightful and a major step forward in the exploration of team processes, this method has yielded conflicting results (Orasanu, 1990; Thornton, 1992). While a majority of research has found that more communication is associated with better performing teams (Foushee & Manos, 1981; Mosier & Chidester, 1991; Orasanu, 1990), subsequent studies have not confirmed these results. For example, Jentsch and colleagues (1995) found that while teams that communicated more identified problems more quickly, they did not necessarily solve problems more rapidly. In addition, Thornton (1992) found

results that indicated that teams who used more communication indicative of situational awareness made more errors, suggesting that the teams were making additional communications to compensate for past mistakes. These conflicting results led researchers to pursue additional communication analysis techniques to further examine the "more is better" concept of team communication.

In addition, just how much team members talk is not the only important aspect of team communication. For example, team members who talk off-task or do not address questions asked by other team members add to the frequency of communication without actually doing anything to help the team. Therefore, 'what' members talk about also has a significant impact on performance. For example, Smith-Jenstch, Zeisig, Acton, and McPherson (1998) developed Team Dimensional Training (TDT), a training system for teams. Out of this development, a communication coding system for classifying teamwork behaviors was developed. The TDT training system used communication coding to evaluate training effectives through the systematic identification of teamwork behaviors that are essential for good team performance. TDT communication categories are structured around four teamwork dimensions: (1) information exchange, which involves knowing what information to pass to team members and when to pass the information, (2) communication delivery, which focuses on how the information is delivered, (3) backup behavior, which involves team members taking action to help alleviate some of the workload on other members of the team, and (4) initiative/leadership, which involves behaviors that help to guide or provide direction for the team (see Table 5). In addition, within these dimensions, sub-dimensions were developed targeting more specific behaviors (see Table 5). These four dimensions were developed by comparing the behaviors of experienced teams with those of

inexperienced teams within the Navy (Smith-Jentsch & Zeisig, 1998). Building on these analyses for the current study, we have two hypotheses regarding team types and teamwork (as defined by Smith-Jenstch & Zeisig, 1998).

Hypothesis 1a: Teamwork dimensions (TDT) will be predictive of team performance in all teams.

We predict, as the literature has said, that all teams need information exchange, support behavior, and leadership to perform effectively. There are certain fundamental needs of all teams, despite interdependency or distribution that must be present to perform effectively. We believe that these dimensions are global so will apply to all team types.

Hypothesis 1b: The manner in which teamwork is implemented, teamwork sub-dimensions, will differ between team types.

The teamwork dimensions identified in TDT are carried out through other, more specific behaviors. These make up the sub-dimensions identified by Smith-Jenstch et al (1998). We propose that while all teams need teamwork (i.e., dimensions), so all effective teams will show teamwork dimensions, we should see differences on what behaviors the different team types use to achieve these teamwork behaviors. For example, the exchange of information will be related to team performance across team type, but will differ based on the sub-dimensions. Due to the nature of these teams and their need for information despite team type, all of these teams will need information exchange teamwork dimensions to perform effectively. However, their needs for certain types of information exchange (i.e., the sub-dimensions) will differ based on their interdependence and their task. Therefore, based on team type, task characteristics, and interdependency, the importance of the sub-dimensions of

information exchange, supportive behavior, and leadership will differ across the 3 team types.

Hypothesis 2: Team type will moderate the strength of the relationship between teamwork and performance.

For example, we predict that leadership will have a stronger relationship to performance in distributed teams. We expect to find stronger relationships between leadership categories and distributed teams, than between leadership and co-located teams. Due to their configuration, distributed teams should benefit more when these leadership categories are communicated and show more deficiencies if leadership behaviors are not communicated.

Communication Sequences. But again, this technique only taps one aspect of team communication: what is said by team members. But it is not only important what is said—or how often it is said—but how (the order) in which it is said. Specifically, similar to the TDT example above, communication can be coded as to what the content is (e.g., action, enemy) and based on that, what order they come in within the communication flow. For this end, the classification has a natural flow, such as a question followed by an answer. In their work with communication analysis, Bowers, Jentsch, Salas, and Braun (1998) applied elements of ESDA, a technique often used by researchers in other domains (Sanderson & Fisher, 1997), to the analysis of *communication sequences*. ESDA, as applied to communication analysis, examines the sequence of communication categories (See Table 6) and closed and open loop communications. In team communication, an important KSA is closed loop communication. Essentially, in order to communicate effectively, team members must "close the loop" in communication. This means that if a team member is asked a question he or she should

answer it or if a team member is given a command, he or she should acknowledge it. This way, all team members are on the same page. These sequences are assessed based on whether or not they occur more often than would be expected by chance. Results supported a general improvement over traditional communication analysis, gleaning information not available through frequency counts, but left a need to examine the generalizability of results across teams. This type of analysis can also clarify confusion on what is being said by looking at the communication surrounding a statement and the context in which it is said.

For example, in past communication coding done by the authors, there came a point where some confusion over how to classify a statement led to the discussion of the importance of sequencing. While coding dyad communications, researchers encountered a phrase that was ambiguous in meaning unless the order of statements was examined. Specifically, a team member would say “there were two of them”. That statement is extremely vague when analyzed by itself. Without an examination of context, the researchers were unable to determine if the statement was task relevant or not. The researchers went back to the preceding statements made by team members and found that the preceding statement was a question (i.e., an uncertainty). The first team member asked “how many enemies do you see?” and the other team member responded “there are two of them”. After examination of the sequence of communication between team members this initially vague statement was not only determined to be task relevant, but was also important for the completion of the task. Similarly, if the preceding statement in the sequence was nontask-related, it would have likely indicated that the statement “there were two of them” was unrelated to the team’s performance and would have just added to the amount of communication.

Based on this, we have made four more hypotheses (Hypotheses 3-6) regarding closed loop communication, sequencing, and team performance.

Hypothesis 3: The amount of communication used will vary based on team type.

We predict that the overall amount of communication will differ based on the task interdependency, distribution, and shared visualization of the team. For example, geographically distributed teams will need to communicate more to complete their tasks. Therefore, co-located teams will have less communication.

Hypothesis 4: Closed loop communication sequences will be significantly related to effective team performance across team type.

Specifically, we predict that closed loop communication will be positively related to team effectiveness, as defined by the authors. In all three team types, we expect to see that effective teams will "close the loop" in communication more often than poor performing teams, replicating Bowers et al. (1998) findings. This means that communication between team members in effective teams will more often consist of open communications (e.g., uncertainty/question) followed by a closed communication (e.g., response) than in ineffective teams. We believe that teams who perform effectively will show more complete patterns in communication.

Hypothesis 5: Open loop communication sequences will be significantly related to ineffective team performance across team type.

Specifically, we predict that open loop communication will be positively related to poor team performance. In all three team types, we expect to see that ineffective teams will use open loop communication sequences more often than more effective teams. Ineffective teams will have more open communication sequences consisting of open communications

(e.g., uncertainty/question) followed by open communications (e.g., uncertainty/question) or closed communications (e.g., responses) followed by open communications (e.g., uncertainty/questions). We believe this based on both the literature and our knowledge of the way effective teams operate. To develop shared knowledge, necessary for effective team performance, teams must use closed loop communication, which enables team members to stay up-to-date and have knowledge of what other team members see and do. By not closing the loop (i.e., open loop communication), team members can become confused and develop divergent knowledge structures leading to ineffective performance.

Hypothesis 6: Team type will moderate the strength of the relationship between closed loop communication sequences and performance.

Specifically, we predict that team type (i.e., the three team types represented by the testbeds based on interdependence level and distribution) will moderate the strength of the relationship between closed loop communication sequences and performance. We believe that all teams need closed loop communication, that it is an essential KSA. However, the amount of closed loop communication sequences necessary to effectively perform will differ based on team interdependency, shared visualization, co-location versus distribution, and shared versus distributed expertise. For example, teams that have distributed expertise, are distributed physically, have no shared visualization, and moderate interdependency will need to close the loop more often than co-located teams with shared expertise, high shred visualization, and high interdependency, since they can coordinate implicitly and communicate nonverbally. Following the rationale depicted in taxonomic literature on team type, it was our belief that teams having different performance arrangements (i.e., differing

levels of co-location and expertise) and interdependencies would see differential patterns of communication.

Hypothesis 7: Team type will moderate the strength of the relationship between open loop communication sequences and performance.

We predict that team type (i.e., the three team types represented by the testbeds based on interdependence level and distribution) will also moderate the relationship between open loop communication sequences and performance in ineffective teams. While all teams are negatively effected by an inability to utilize closed loop communication, the characteristics of the team (e.g., distribution) will moderate how much of an effect open loop communications will have on their overall performance. For example, co-located teams with high interdependency can depend on other factors (e.g., nonverbal communication and environmental cues) to close the loops. In addition, less interdependent team performance is dictated less by the ability to close the loop.

Two areas that the current study attempts to address are 1) differences across team type and 2) team communication and performance.

METHOD

These team configurations took the form of three distinct action team types that attempted to replicate real world teams. Specifically, we chose real life military teams to serve as guides for external validity in the development of our teams and testbeds: combat teams, command-and-control teams, and recon teams. Therefore, the three teams developed by researchers are 1) the co-located high visualization team, 2) the distributed low visualization team, and 3) the distributed no visualization team. The co-located high visualization team was designed to replicate a combat team for military operations. This team

is physically co-located, has some distributed expertise and some shared expertise, a visual space that is common (i.e., high shared visualization), and, as a result, has team interdependency. The distributed low visualization team was designed to replicate command-and-control teams, resulting in physical distribution (i.e., team members are not face-to-face), distributed expertise (e.g., team members have different roles or skills), low shared visualization (i.e., team members see low fidelity screens that look the same), and reciprocal interdependence. The distributed low visualization team was designed to replicate military reconnaissance teams. This team has a distributed physical space , distributed expertise, no shared visualization (i.e., team members see completely different things), and sequential interdependence. We believe that these three teams are examples of real world teams and have characteristics that can affect team process and performance differentially.

Participants

Participants were 152 undergraduate students (76 two-person teams) at a large southeastern university that were randomly assigned to teams and conditions. Participants had a mean age of 21.28 (sd=4.38), with age ranging from 17 to 49. Participants were recruited through the university by flyers, class visits, and a university sponsored web site. Participants were compensated with 1 extra credit point per half hour of participation, which was submitted to their professor electronically.

Design

The design was a between subjects design with one factor, interdependence, which consisted of 3 levels: team interdependency, reciprocal interdependency, and sequential interdependency. Each condition was created by manipulating key factors that would create a task environment that required a different level of interdependency. These factors include

distribution of expertise, physical distribution, and shared visualization. These factors drove the requirements of the task and made varying levels of interdependency possible. For a full breakdown of factors by condition, see Table 7.

Equipment

The equipment used for the present study varied by condition. In the team interdependent/Ghost Recon condition, three computers were used: one computer for each team member to practice during training and to play the mission; one computer for training using Microsoft Powerpoint. In the reciprocal interdependent/DDD condition, two computers were used: one for each team member to play the mission. These computers were also used for training using Microsoft Powerpoint. In the sequential interdependent/Operation Flashpoint condition, one computer was used for the team member playing the mission immersed within the game and for training.

The equipment required for data collection and communication capture included headsets, a sound mixer, a television, and a VCR. These devices were attached to the computers in the interdependent/Ghost Recon and sequential interdependent/Operation Flashpoint conditions. In the the reciprocal interdependent/DDD condition, a software program, Snagit by Camtasia, was used to capture the communication and screen shots, due to the low fidelity interface. Snagit was not used for the other two conditions due to the heavy graphical load of the Ghost Recon and Operation Flashpoint games.

Manipulations

Team Interdependency Condition: Ghost Recon. Ghost Recon produced by Red Storm Entertainment is a team-based first-person shooter game that provides realistic simulations of virtually co-located combat teams. For this study, an existing mission within

the game was modified in order that the objectives given to the team and the resources available to each team member required coordination at a team level of interdependence. Teams were given task specific training, including basic movements and controls within the game as well as coordination strategies. For example, teams were taught to work together to clear a house (i.e. the team member without grenades would open the door to the house for his/her teammate to throw a grenade inside). The team training lasted approximately 2 hours and was followed by the experimental mission.

Within the mission, teams were given four objectives: Objective 1-Clear House 1, Objective 2-Clear House 2, Objective 3-Destory SAM Site, Objective 4-Reach Extraction Zone with Team intact. In order to complete these objectives one team member was given grenades to use to clear the houses, while the other team member was given an M136 rocket launcher used to destroy the SAM site, in addition to their M4 rifles. The different tools given to each member created both shared and distributed expertise. Participants were also taught different strategies, utilizing different tools, based on expected presence of friendlies, enemies, and unknowns. Both members of the team saw a graphic environment of the same town with computer representations of the other team member, showing their movement and position within the game. These elements created high shared visualization, as described above. Performance in the task was measured by a team's ability to complete the objectives and the time to complete each objective.

Reciprocal Interdependency Condition: DDD (Dynamic Distributed Decision Making) is a distributed team-based synthetic task simulation developed by Aptima, Inc. used to study how effective teams interact in complex environments. The task within the simulation best represents a command-and-control task with reciprocal interdependence.

Specifically, teams in this condition were tasked with correctly identifying and destroying enemy targets. Each team member had a view of the same screen, creating shared visualization, but the fidelity was low and the information about the other team member actions and tasks was limited (i.e., the screen showed the results of there actions on a radar screen), making the shared visualization low. A one hour training session was provided to each team prior to the 30 minute mission. Participants received training on the basic purpose of the DDD simulation as well as task specific training on the controls used in the simulation and the different methods of communication (i.e. communication within the DDD simulation can be either verbal or electronic).

Within each team, one participant had access to information on whether a target was a friendly or an enemy. The other team member was tasked with destroying enemies based on the information given to them by the other team member. This design created both reciprocal interdependence (i.e., information had to be exchanged back and forth to complete the task) and distributed expertise (i.e., each team member had differentiated tasks and abilities within the game). The information had to be exchanged in a timely manner so the other person on the team would have time to take appropriate action (e.g. destroy target if it was an enemy). A total of 25 enemy targets were presented during the mission. Points were deducted from the team score for friendly targets destroyed or for enemy targets not destroyed.

Sequential Interdependency Condition: Operation Flashpoint. Operation Flashpoint published by Codemasters was modified so that teams were tasked with gathering and verifying information from a town. Although Operation Flashpoint is a first-person shooter game, it was modified so that the participants had no ammunition and there were no enemies present in the scenario. Teams were given a brief training session of approximately 30

minutes to become familiar with the basic movements required to complete the task and with the structure of the map and statements. The training also included instruction on how to use the computer-based tools (e.g. the compass) needed to successfully complete the mission.

In order to simulate a distributed team, one participant was immersed in the gaming environment and tasked with navigating through the town, while his/her teammate was given a list of statements, grouped as eight sets of three, gathered from eight enemy fighters that had been taken hostage in raids prior to the mission. In order to complete the task, participants relied on 7" x 7" physical maps of the town, the computer-based compass, and other computer-based tools, to verify the truthfulness of written statements by identifying locations and objects in the virtual environment. Each team member had a distinct set of tasks and responsibilities creating distributed expertise. Because the team member who was outside the game had to feed the team member in the game information the team member in the game could not start his task until the team member outside of the game completed their task (i.e., found the item to look for and it's location on the map, then give appropriate directions to the other team member), creating sequential interdependence. The team members had no shared visual field. The number of correctly verified statements, such as a whether or not a particular object was found in its stated location (circled by participants as "true" or "false" during the task), served as the outcome measure. The maximum number of points available during the 30 minute performance segment was 24.

Materials/Measures

Within the study, surveys were administered across testbeds to tap demographic information. The demographic questionnaires collected general information about participants. These questions addressed computer use, gaming experience, military

experience, age and gender. This information was used during analysis to control for computer skill effects on performance.

In addition to the performance measures discussed within the design section and the surveys, the communication was analyzed for all team types based on interdependency (i.e., team, reciprocal, and sequential interdependent teams). Communication was analyzed for content as well as sequence of communication. Content analysis was based on teamwork categories and techniques proposed by Smith-Jenstch et al. (1996), team dimensional training (TDT), and sequence analysis was based on the categories and techniques used by Bowers et al. (1998), the Exploratory Sequential Data Analysis (ESDA).

TDT systematically identifies teamwork behaviors that are essential for good team performance, through communication (e.g., verbal and nonverbal). TDT utilizes four teamwork communication categories: information exchange, communication delivery, support behaviors, and leadership. We have excluded the communication delivery category, as defined by TDT, from our analyses because we believe there will be no relationship between communication delivery categories and performance in any of the teams. This is not to say that communication is not important, but that communication as defined in the TDT refers to specific communication (e.g., SOP) that were not taught to the subjects and, therefore, not expected in the analysis. ESDA, discussed earlier, was also used to assess closed loop communication and sequencing. Communication coders were trained until they reached an acceptable interrater reliability over 80%.

Procedure

Participants were first asked to sign an Informed Consent form, then were randomly assigned to a roles, teams, and conditions. The participants next filled out demographic

questionnaires including questions about age, gender, computer use, and military experience. In addition to providing information about participation, the portion addressing computer use and skill was later used in the analyses to control for any disparate computer expertise of participants. Participants were then given training relevant to their task (e.g., how to use the weapons, identify targets, how to move within games). Following training, which included practice sessions, participants were given the Dundee Stress State Inventory (trait). After a short break, teams were asked to complete a 30 minute performance mission. Following completion of the experiment, participants were debriefed.

RESULTS

Due to difficulties with the technology used to record the communication, several teams were excluded from the analyses because the communication was inaudible. Therefore, for the current analyses (i.e., TDT, ESDA) represent a total of 52 teams, 16 teams per interdependence level (N=104). Table 8 provides intercorrelations between performance, team type and the TDT teamwork dimensions. Table 9 provides intercorrelations between performance, team type, and the communication categories utilized in the sequential analyses.

In order to investigate the relationships between TDT teamwork communication behaviors and performance in addressing hypothesis 1a, a linear regression was conducted across teams despite interdependence level for the dimensions. The analysis revealed that the dimensions for communication of teamwork behavior were predictive of performance, predicting 41% of the variance, $F(44, 3) = 10.11$, $p = 000$, with support behaviors ($\beta = .56$, $p = .002$) and leadership ($\beta = .13$, $p = .001$) demonstrating significant effects on performance.

While this further supports an overall relationship between teamwork dimensions and performance, the tasks themselves may require fewer or more communications, so simply using frequency of such dimensions can distort the results. Therefore, we also calculated the percentage of dimensions and sub-dimensions based on overall communication of participants within a condition to ensure that communication is not influenced by the varying tasks. We again conducted a regression analysis with the dimensions, this time using the percentages. We found that the percentages of the dimensions were predictive of team performance, predicting 42% of the variance, $F(45, 2) = 16.42$, $p = .000$, with information exchange ($\beta = -.61.06$, $p = .000$), support behaviors ($\beta = 208.326$, $p = .001$), and leadership ($\beta = 54.71$, $p = .000$) demonstrating significant effects on performance.

While this supports an overall relationship between teamwork dimensions and performance, we would predict their may be differences when examined at a deeper level, namely the sub-dimensions in hypothesis 1b (i.e., teamwork sub-dimensions, as examined through communication, will differ between team types). Therefore, to further investigate the effect of team type on performance, we ran separate linear regressions for each of the three testbeds with performance as the dependent variable and the sub-dimensions of the TDT teamwork behaviors as predictor variables. These results gave further insight to the moderating effect of team type on performance. For both the team interdependence and the reciprocal interdependence conditions, none of the TDT sub-dimensions significantly predicted performance indicating that the sub-dimensions did not offer a unique contribution to the prediction of performance. Instead, it is the combination of the individual sub-dimensions that is significantly related to performance in the team interdependence condition. For the sequential interdependence condition, two of the information exchange

sub-dimensions gave unique and significant contributions to the prediction of performance: seeking information from all sources ($\beta = .581$, $p = .032$) and passing information before being asked ($\beta = .968$, $p = .014$).

In order to address hypothesis 2 (i.e., team type will moderate the strength of the relationship between TDT teamwork communication dimensions and performance), a regression analysis was run based on the interaction between interdependence team type and communication. While the predictive power of communication alone was significant at 42%, when the interaction between communication and interdependency level (e.g., team) were included, these interactions predicted 64% of the variance in performance, $F (45, 2) = 39.79$, $p = .000$, ($\beta = .028$, $p = .001$).

In order to test hypothesis 3 (percentage of overall communication will vary by team type) a 3 x 8 ANOVA was performed between team type and the 8 communication categories outlined by Bowers et al. (1998): plan, action, acknowledgement, response, uncertainty, fact, non-task related communications, and communication with the experimenter. A percentage of overall communication for each team type for each category communication was used in the analysis in order to compare across teams types. Results supported a difference based on team type on several of the categories. Action communications (open), $\underline{F} (45, 2) = 41.46$, $p = .000$, acknowledgement communications (closed), $\underline{F} (45, 2) = 16.38$, $p = .000$, responses (closed), $\underline{F} (45, 2) = 29.18$, $p = .000$, uncertainty communication (open), $\underline{F} (45, 2) = 45.52$, $p = .000$, and facts (open), $\underline{F} (45, 2) = 65.87$, $p = .000$, were all significantly different based on team type. Tukey's post hoc tests revealed that reciprocal interdependence teams used action and acknowledgement communications less than the team and sequential interdependent teams, $p < .05$. All three

team types used significantly different percentages of responses, uncertainties, and facts, $p < .05$. For responses, the reciprocal interdependence teams had the lowest mean percentage ($m = 5.41$, $sd = 4.14$), the team interdependent team were next ($m = 8.48$, $sd = 2.78$), and the sequential interdependent teams had the highest mean percentage of responses ($m = 14.00$, $sd = 2.49$, $p < .05$). Uncertainties followed the same pattern: the reciprocal interdependence teams had the lowest mean percentage ($m = 10.11$, $sd = 7.53$), the team interdependent team were next ($m = 20.13$, $sd = 4.82$), and the sequential interdependent teams had the highest mean responses ($m = 28.16$, $sd = 3.52$, $p < .05$). For facts, the sequential interdependent teams again had the highest mean responses ($m = 74.58$, $sd = 16.31$, $p < .05$), the team interdependent team were next ($m = 42.83$, $sd = 12.68$), and the reciprocal interdependence teams had the lowest mean percentage ($m = 24.24$, $sd = 6.75$).

To address our fourth hypothesis (i.e., closed loop communication will be significantly related to effective team performance across team type), we did a median split and divided teams into high performing and low performing teams. We then examined the z scores for the parameter estimates. However, our hypothesis was not supported. Across team types, high performing teams only used closed loop communication 18% of their total communication sequences ($\underline{z} = 0.07$), while poor performing teams only used closed loop communication 18.7% of the time ($\underline{z} = 0.19$). However, when examined at the team type level, two of the team types had slightly more closed loop communication sequences for high performing teams. In the reciprocal interdependency team (i.e., DDD), high performing teams used closed loop communication 12.8 % of their total communication ($\underline{z} = 12.80$), while poor performing teams in this condition only used closed loop communication 9.3% of the time ($\underline{z} = -1.22$). Teams in the sequential interdependency condition(i.e., Operation

Flashpoint) who performed more effectively used closed loop communication 25.9 % of the time ($\underline{z} = 1.28$), while poor performing teams used closed loop communication only 24.6 % of the time ($z = 1.09$). In the team interdependency condition (i.e., Ghost Recon), poor performing teams used closed loop communication 17.5 % of the time ($\underline{z} = 0.02$), while high performing teams used closed loop communication 14.2 % of the time ($\underline{z} = -0.48$). Conversely, our fifth hypothesis was not supported either (i.e., open loop communication sequences will be significantly related to ineffective team performance across team type).

To test Hypothesis 6 and 7 (team type, as defined by the authors, will moderate the strength of the relationship between closed/open loop communication sequences and performance), we first performed a regression analysis with performance as the dependent variable and the interaction between interdependent team type and communication as the predictor variable. We found that this interaction predicted 62% of the variance in performance ($\beta = 1.31$, $p = .002$), supporting our hypothesis.

Next, we performed a regression analysis with performance as the dependent variable and open loop communication, level of interdependency (team type), and the interaction of open loop communication and level of interdependency as predictor variables. Results indicated no significant effects for open loop communication by itself but a significant interaction between open loop communication and level of interdependency (team type) on performance. Specifically, the interaction accounted for 42.4% of variance in performance. Thus Hypothesis 6 was supported for both the closed loop and open loop communication predictions.

DISCUSSION

Are all teams created equal? The results from the current study indicate no. There are differences in how teams communicate, what they say, and how often they say things based on their level of interdependency, distribution (expertise and geographical), and the amount of cues they receive in the environment. While this question can not be emphatically answered based on one study, these results at least get at the question of treating all teams the same. Training, evaluating, and using all types of teams in the same way, despite level of interdependence or distribution, may be a disservice to the science and practice of teams. The complexity of teams and teamwork demands examination at a deeper level.

The purpose of this study was two fold: 1) to determine if different types of teams exhibit different team processes in order to complete team goals; 2) to replicate and extend communication analysis research across and between team type. In order to accomplish this, the researchers developed a testbed that was representative of 3 types of interdependency (i.e., team, reciprocal, and sequential). The results of the current study offer preliminary support for differences in communication between teams with different levels of task interdependency. Although there was not complete support for all hypotheses, the evidence presented here further informs the general knowledge base of teams and communication sequencing and behaviors. Specifically, these results lend support to differences in communication based on team type when classifying teams on several variables. This is especially important because, as mentioned in the introduction, researchers and practitioners alike typically treat a team as a team, with little or no distinction between these variables. If, in fact, teams communicate and, therefore, process differently, teams must be trained, evaluated, and assigned differently.

The relationship between performance and communication was supported by our first few analyses, using both the raw number of communications and percentages of overall communications. While it was surprising that only 2 overall communication teamwork dimensions (i.e., support behavior, leadership behavior) were significantly related to performance, examination at a deeper level (i.e., on the sub-dimensions) showed more significant relationships between teamwork behaviors communicated and performance across the teams. Within all three dimensions examined in the current study, sub-dimensions within each were significantly related to performance, supporting the importance for at least aspects of information exchange, supporting behaviors, and leadership to team performance regardless of team type.

Regarding the moderating effect of team type, the regression showed a significant amount of the variance (i.e., 64%) of overall performance explained by the interaction of team type and communication, significantly more than the 42% variance explained by communication alone. This supports our hypothesis that the type of team will significantly affect the relationship between their communication and performance. This is perhaps one of the most interesting findings in that it lends support to the notion that not all teams are alike, since they have different communication requirements to perform a task effectively.

Another surprising finding was the failure to support the presence of more closed loop communication sequences in high performing teams across team type. While low performing teams used closed loop communication only slightly more (i.e., 18.7% of their communication compared to 18% in high performing teams), the trend is indicative of the need for more research in this area, since conflicting results continue to plague the literature.

Upon examining the closed loop communication between team type, however, an explanation seems possible. Although no significant differences were found, two of the three team types did exhibit slightly more closed loop communications in high performing teams. One of those teams, the reciprocal interdependence team utilizing the DDD tool, is indicative of the type of team usually used in team research, a command-and-control team. This type of team exhibiting more closed loop communications among high performing teams supports previous findings of the relationship between closed loop communication and effective performance (Bowers et al., 1998). While this was a promising result, the fact that not all team types followed this same communication pattern emphasizes the need for further research in the area of team communication within different team types. One possible explanation as to why the team interdependence teams did not use closed loop communication is that they had a higher shared visualization, making them more capable of 'closing the loop' nonverbally. The nonverbal communication of co-located teams should be further explored and is a weakness of the current study since it was not included in the analyses.

The moderating effect of team type on performance was again supported with the ESDA analyses for both open and closed loop communication. Furthermore, the differences between categories by team type illustrate the different types of communications these teams needed to accomplish their task. The different types of team communications are indicative to different processes that may be occurring during teamwork. While all teams require the KSAs demonstrated by successful teams, this project lends credence to the theory that different processes may be more or less important for different types of teams in successfully completing their tasks.

Limitations and Future Research

Like many studies, our investigation has some limitations. First, the sample population consisted of students and was collected in the lab. Future studies into the equality of different teams would benefit from more in depth investigation in the field and through examination of intact, professional teams. The ad hoc teams utilized in the current study was only the first step in investigating such a complex phenomena. Furthermore, our team types, based mostly on task interdependencies, have other characteristics that should be parsed out in future research. For example, while the reciprocal and sequential teams have varying levels of interdependence, they both are distributed but have varying levels of shared visualization and distributed expertise. Based on this, we can not say that their task interdependency was the only variable that led to varying levels of communication definitively. While there are bound to be many confounds in looking at different types of teams, some of the more salient characteristics can be examined to determine what characteristics lead to different communication levels and sequencing. Finally, a major limitation of the current study was the attrition of the sample population. Much of the loss of teams (32% attrition) can be blamed on technological difficulty, but future studies need to be designed in a way that communication can be more accurately and effectively captured.

Applications

The implications of this research are widespread. For example, based on our results, it may not be appropriate to apply a general team finding (e.g., all teams use more closed loop communication to perform effectively) based on research using one team type (e.g., command and control) to another team type (e.g., a recon army team). The results reported here support further examination of teams, team training, and how we define teams. If there

are not the overarching rules for effective teams that we have always relied on when it comes to communication, researchers and practitioners must more carefully define the teams they use and must move beyond the general team training typically applied regardless of team type. Ultimately, although further research is needed, potential applications of this research include the design of team training based upon specific team characteristics and structure. As researchers and practitioners, we must set boundaries for conditions around theories and remember that context is very important. This study begins to do that by explicitly defining the teams and their characteristics to ground theories of team performance and communication in context. Future research must examine both team models of effectiveness and team training to determine their applicability across team type. The findings based on communication analysis lead us to question: Is there a 'team g' (i.e., a general team type that applies to all teams) or do we differentiate team type based on specific characteristics (e.g., interdependency)?

REFERENCES

Argote, L. & McGrath, J. E. (1993). Group processes in organizations: Continuity and change. In C. Cooper and I. T Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 333-389). Chichester, England: Wiley.

Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors, 40(4),* 672-679.

Brannick, M., Roach, R., & Salas, E. (1993). Understanding team performance: A multimethod study. *Human Performance, 6*, 287-308.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo, E. Salas, & Associates (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco, CA: Jossey-Bass.

Devine, D. J. (2002). A review and integration of classification systems relevant to teams in organizations. *Group Dynamics: Theory, Research, and Practice, 6(4),* 291-310.

Driskell, J. E., & Salas, E. (1992). Collective behavior and team performance. *Human Factors, 34,* 277-288.

Hackman, J. R. (Ed.) (1990). *Groups that work (and those that don't): Creating conditions for effective teamwork* (pp. 315-342). New York: Prentice-Hall.

Hackman, J. R., and Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: a review and proposed integration. In: Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology*, vol. 8. New York: Academic Press.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributive expertise. *Journal of Applied Psychology, 80*(2), 292-316.

Hollenbeck, J.R., Sego, D.J., Ilgen, D.R., Major, D.A., Hedlund, J., and Phillips, J. (1997). Team decision making accuracy under difficult conditions: Construct validation of potential manipulations and measures using the TIDE² simulation. In M. Brannick, E. Salas and C. Prince (eds.), New Directions in Team Measurement, pp. 111-136. San Francisco: Jossey Bass.

Kleinman, D., Pattipati, K. Luh, P., and Serfaty, D. (1992). Mathematical models of team performance: A distributed decision-making approach. In R. Swezey and E. Salas, (Eds.), Teams: Their Training and Performance (177-218). Norwood, NJ: Ablex Publishing Company.

Kozlowski, SWJ (1996, March). TEAMS / TANDEM: An experimental platform for examining skill acquisition, adaptability, and effectiveness at individual and team levels of analysis. Presented at the Personnel Human Resource Research Group, University of Florida, Gainesville, FL.

Matthews, G., Falconer, S., & Campbell, S. (April 1997). *Assessment of Task Induced Stress.* Paper presented at BPS conference. Edinburgh, Scotland.

Morgan, B. B., Jr., Salas, E., & Glickman, A. S. (1993). An analysis of team evolution and maturation. *Journal of General Psychology, 120*(3), 277-291.

Orasanu, J. (1990). Shared Mental Models and Crew Decision Making. Cognitive Science Laboratory Report #46. Princeton, NJ: Princeton University.

Pintrich, P.R., Smith, D.A.F., Garcia, T., & McKeachie, W.J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53, 801-813.

Saavedra. R., Earley, P. C., and Van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology, 78(1),* 61-72.

Salas, E., & Cannon-Bowers, J. A. (2000). Teams in organizations: Lessons from history. In M. M, Beyerlein (Ed.), *Work teams: Past, present and future* (pp. 323-331). Netherlands: Kluwer Academic Publishers.

Sanderson, P.M. and Fisher, C. 1994. Exploratory Sequential Data Analysis: Foundations. *Human-Computer Interaction*, 9, 251-317.

Smith-Jentsch, K.A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress: Implications for Individual and Team Training* (271-297), Washington, D.C.: American Psychological Association.

Steiner, I. (1972). *Group process and productivity.* New York: Academic Press.

Stevens, M. J. & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management, 20*, 503-530.

Sundstrom, E. D. (1999). *Supporting work team effectiveness: Best management practices for fostering high performance.* San Francisco, CA: Jossey-Bass, Inc.

Sundstrom, E. McIntyre, M., Halfhill, T., and Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics, 4(1),* 41-67.

Thompson, J. D. (1967). *Organizations in action.* New York: McGraw-Hill.

Thornton, R. C. (1992). *The effects of automation and task difficulty on crew coordination, workload, and performance.* Unpublished doctoral dissertation, Old Dominion University, Norfolk, VA.

Van de Ven, A.. Delbecq A., & Koenig, R. Jr. (1976). "Determinants of coordination modes within organizations." *American Sociological Review 41*, 322-338.

Table 1. Team Types (Sundstrom et al., 2000)

Team Type	**Definitions**
Action	Groups that conduct complex, time-limited performance events that involve audiences, challenging environments, or adversaries
Project	Teams that carry out projects that are time-limited, specialized, and defined; typically teams are cross-functioning and disband after finishing a project
Production	Teams who produce tangible products repeatedly and are on the frontline
Service	Teams that conduct transactions repeatedly with customers
Management	Teams that coordinate work of subordinates through budgeting, planning, policy making

Table 2. 6 TYPES OF WORK GROUPS (Sundstrom, et al. 2000)

Team Type	**Characteristics**	**Examples**
Production	- *Work cycles typically are repeated, continuous processes that may be shorter than team's life span* - Can be autonomous or semi-autonomous - Low differentiation - High integration	- Assembly teams - Processing groups - Manufacturing crews
Service	- *Employees cooperate to conduct repeated transactions with customers* - Can also be autonomous or semi-autonomous - Low differentiation - High integration	- Sales teams - Maintenance groups - Airline attendant teams
Executive/Management	- *Consist of a Sr. manager and those managers/supervisors who report to him or her.* - Concerned with: joint planning, budgeting, logistics, - High integration - Moderate differentiation	- Corporate executive teams - Regional steering committees
Project Teams (Task Forces)	- *Carry out defined, specialized, time-limited projects and disband upon completion* - Cross-functional - High differentiation - Low integration	- Research groups - Planning teams - Development teams - Task forces
Action & Performing	- *Conduct complex, time-limited events with adversaries, audiences, or challenging environments*. - High integration - High differentiation	- Musician groups - Firefighters - Negotiating teams - Surgery teams - Cockpit crews
Advisory/Parallel	- *Temporary assembly of persons grouped to solve problems and recommend solutions.* - Work outside of, and parallel with production processes - Varied differentiation - Low integration	- Quality circles - Employee involvement groups - Selection committees

Table 3: Taxonomy of Workgroups (Devine, 2002)

Workgroup type	Function	Exemplars	Fundamental work cycle	Physical abilities	Temporal duration	Task structure	Active resistance	Hardware dependence	Health risk
Executive	• Plan • Direct	• Board of directors • City council	Decision/plan	Low	Indefinite	Low	No	Low to moderate	Low
Command	• Integrate • Coordinate	• Control tower • Combat center	Order	Low	Shift	High	No	High	High
Negotiation	• Deal • Persuade	• Labor–management • International treaty	Deal/agreement	Low	Mission	High	Yes	Low	Low
Commission	• Choose • Investigate	• Search committee • Jury	Decision/judgment	Low	Mission	Low to moderate	No	Low	Low
Design	• Create • Develop	• R&D • Marketing	Model/prototype	Low to moderate	Mission	Low	No	Moderate	Low
Advisory	• Diagnose • Suggest	• Quality circle • Steering committee	Report/plan	Low	Mission	Moderate to high	No	Low	Low
Service	• Provide • Repair	• Fast food • Auto service	Customer	Moderate to high	Shift	Moderate to high	No	High	Low
Production	• Build • Assemble	• Home construction • Automotive assembly	Unit	High	Indefinite	High	No	High	Moderate
Performance	• Enact • Display	• Movie cast • Orchestra	Scene/routine	High	Indefinite	High	No	Low	Low
Medical	• Treat • Heal	• Surgery • ER	Patient	High	Shift/mission	High	No	High	High
Response	• Protect • Rescue	• Fire Station • Paramedic	Call/run	High	Shift	Low	No	Moderate	High
Military	• Neutralize • Protect	• Infantry squad • Tank crew	Order	High	Indefinite	Low to moderate	Yes	High	High
Transportation	• Convey • Haul	• Airline cockpit • Train crew	Trip/leg	Moderate to high	Shift	High	No	High	High
Sports	• Compete • Win	• Baseball • Soccer	Play/possession	High	Indefinite	Low	Yes	Low	Moderate

Note. R&D = research and development; ER = emergency room.

Table 4. Task Interdependency (Saavedra et al., 1993)

Task Interdependency	**Definitions**
Pooled	Each member makes a contribution to the output without the need for direct interaction
Sequential	Team members have different roles and perform different tasks in a prescribed order; One team member must act before the other can
Reciprocal	Team members have different roles and expertise; use temporally lagged 2 way interactions
Team	Team members jointly diagnose, problem solve, and collaborate; it is up to group discretion to decide the particular course that they take to achieve outputs/goals

Table 5: Categorization of TDT Teamwork Behaviors

Information Exchange	**Communication Delivery**
• Utilizing all available sources of information • Passing information to the right persons without having to be asked • Providing 'big picture' updates	• Proper phraseology • Completeness of standard reports • Brevity/avoiding excess chatter • Clarity/avoiding inaudible communications
Supporting Behavior	**Initiative/Leadership**
• Monitoring & correcting errors • Providing & requesting backup or assistance to balance	• Providing guidance or suggestions • Stating priorities

Table 6: Communication Categories from Bowers et al. (1998)

Category	**Defined as:**
Plan	A statement that communicates an expected or anticipated plan, event, goal, etc.
Action/ Command	A statement/command that outlines an action (e.g., gives directions)
Acknowledgement	A one-bit statement that confirms the receipt of a command or fact
Response	A statement in response to a plan, action, uncertainty
Uncertainty	A statement of uncertainty expressing a question that seeks information from the other team member
Facts	A statement about the environment or other fact

Table 7: Team Configurations and Characteristics

	Team I: Co-Located High Visualization Team	**Team II: Distributed Low Visualization Team**	**Team III: Distributed No Visualization Team**
Testbed	Ghost Recon	DDD	Operation Flashpoint
Physical Location	Co-Located	Distributed	Distributed
Expertise	Shared and Distributed	Distributed	Distributed
Shared Visual Space	High Shared Visualization	Low Shared Visualization	No Shared Visualization
Interdependency	Team Interdependency	Reciprocal Interdependency	Sequential Interdependency

Table 8: Intercorrelations for TDT Teamwork Behaviors

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Performance		-.83**	.33*	.51**	-.11	.36*	-.13	.43**	.33*	.37**	.50**	.51**	-.26
2. Team Type[1]			-.26	-.62**	.28	-.39**	.18	-.49**	-.27	-.45**	-.55**	-.56**	.28
3. Information Exchange (IE)				.54**	.52**	.62**	-.22	.25	.11	.23	.38**	.38**	-.13
4. IE: Seeking					-.38**	.78**	-.20	.32*	.20	.28	.72**	.74**	-.33*
5. IE: Passing						-.29*	-.07	.06	-.01	.06	-.31*	-.32*	.19
6. IE: Responding							-.16	.10	.05	.09	.60**	.62**	-.26
7. IE: Providing								-.08	-.07	-.07	-.15	-.15	.06
8. Supporting Behavior (SB)									.25	.98**	.06	.06	-.10
9. SB: Correcting										.05	.21	.22	-.11
10. SB: Providing											.01	.02	-.08
11. Leadership (L)												.99**	-.14
12. L: Providing													-.22
13. L: Stating													

[1]Team Interdependent Team=1; Sequential Interdependent Team=2; Reciprocal Interdependent Team=3

* $p < .05$, ** $p < .01$

Table 9: Intercorrelations for Communication Categories

	1	2	3	4	5	6	7	8
1. Performance		-.83**	.43**	.59**	.56**	.52**	.56**	.67**
2. Team Type[1]			-.36*	-.65**	-.63**	-.53**	-.61**	-.70**
3. Plan				.23	.29*	.35*	.23	.36*
4. Action					.70**	.69**	.77**	.83**
5. Acknowledgement						.81**	.87**	.78**
6. Response							.92**	.78**
7. Uncertainty								.87**
8. Total Communication								

[1]Team Interdependent Team=1; Sequential Interdependent Team=2; Reciprocal Interdependent Team=3

* $p < .05$, ** $p < .01$

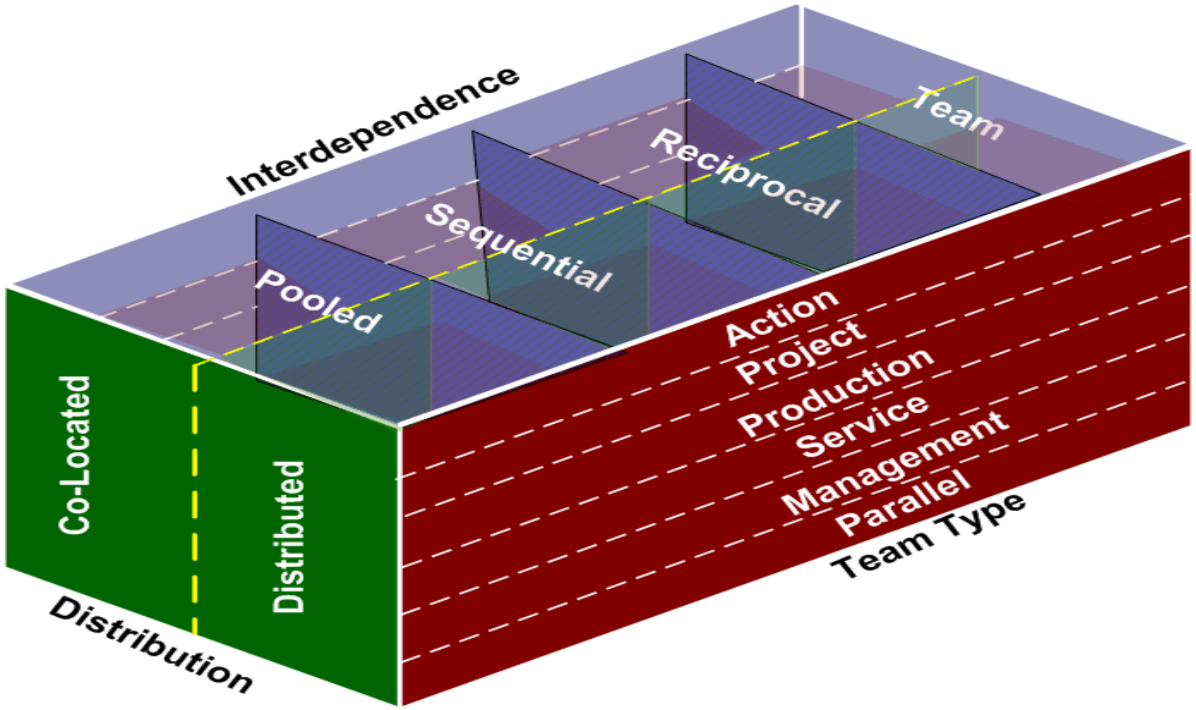


Figure 1. Team Typology Matrix

Heather A. Priest received her M.S. in Experimental Psychology in 2001 from Mississippi State University. She is currently a PhD candidate in the Applied Experimental and Human Factors Psychology program at the University of Central Florida and has been employed as a graduate research assistant at the Institute for Simulation & Training since 2001.

C. Shawn Burke earned her doctorate in Industrial/Organizational Psychology from George Mason University in 2001 and is currently working as a Research Scientist at the University of Central Florida, Institute for Simulation and Training. Her expertise includes teams and their leadership, team adaptability, team training, measurement, evaluation, and team effectiveness.

Joseph W. Guthrie Jr. received his M.S. in Industrial/Organizational Psychology from the University of Tennessee at Chattanooga in 2003. He is a second year PhD student enrolled in the Applied Experimental and Human Factors Psychology program at the University of Central Florida and works as a graduate research assistant at the Institute for Simulation & Training.

Eduardo Salas is Trustee Chair and Professor of Psychology at the University of Central Florida. He also holds an appointment as Program Director for Human Systems Integration Research Department at the Institute for Simulation & Training. Dr. Salas is a Fellow of the American Psychological Association (SIOP and Division 21), the Human Factors and Ergonomics Society. He received his Ph.D. degree (1984) in industrial and organizational psychology from Old Dominion University.

Clint A. Bowers is the Assistant Dean of Research at the University of Central Florida. He received his Ph.D. in Psychology in 1987 from the University of South Florida. He has worked in the area of team performance and training for the past 15 years.

Neuropsychological Assessments in Extreme Environments

Michael Lowe, Ph.D., Wayne Harris, Ph.D., Robert Kane, Ph.D., Louis E. Banderet, Ph.D., Daniel Levinson, Ph.D., , and Dennis L. Reeves, Ph.D.

Abstract

In this paper, we summarize studies that used ANAM tests to assess the effects of environmental stressors. The findings document performance changes resulting from conditions relevant to military operational medicine. These conditions included radiation exposure, toxins, high altitude, undersea conditions, fatigue, and military training.

Humans often function in difficult and demanding environments or under less than optimal conditions. These environments may involve extreme temperatures, sleep deprivation, toxic exposures, or changes in normal ambient pressure which occur when at altitude or engaged in under sea activity. In some cases, extreme environments are a result of deliberate choices such as exploring space or the ocean depths. In other cases, extreme environments result from adverse circumstances, including war or an accident. Measuring the effects of extreme environments on cognition was a motivation for the development of automated, repeatable neurocognitive test batteries within the Armed Forces. In the present article we review data from studies where ANAM tests were used to study the effects of environmental stressors. The review includes studies assessing the effects of radiation, toxins, high altitude, undersea conditions, and high stress military training on performance. The paper combines and summarizes data from peer reviewed journals, technical reports, and meeting presentations.

Assessment of cognitive changes following exposure to radiation and toxins

Radiation

To assess the effects of exposure to varying levels of ionizing radiation on neurocognitive abilities, Gamache and his colleagues (2004) conducted a four-year longitudinal study in and near Chernobyl, Ukraine. Participants were volunteers who resided in Ukraine during and after the Chernobyl Nuclear Power Plant accident that occurred 26 April 1986. From 1995 to 1998, a subset of selected ANAM measures, which had been translated into Ukrainian, was administered annually to a control and three radiation exposure groups. Controls (n=31) were healthy volunteers who resided in Ternopil, 450 km east of Chernobyl. The radiation exposed personnel consisted of three groups: eliminators, forestry workers, and agricultural workers. The eliminators (n=36) were Chernobyl decontamination and reconstruction workers who went into the area immediately and shortly after the disaster; they resided in a hospital in Kiev during the study. Forestry workers (n=29) resided and worked in the Ovruch forest, 25-30 km southwest of Chernobyl. Agricultural workers lived and worked in the village of Razumnitsia, approximately 150 km south of Chernobyl.

Analyses were performed using accuracy and efficiency (throughput) scores for both across-time and between-group comparisons. Correlations were also calculated between neurocognitive performance and levels of radiation, which were obtained for the individuals in the 3 exposure groups. Analyses of 1995 test results (Gamache, 2005) indicated that the eliminators were significantly and globally more impaired on measures

of neurocognitive performance as compared to the Controls. Further, Forestry and Agricultural workers were impaired on many cognitive tests. Analyses of 1996 retest data revealed that neurocognitive performance of personnel in the exposed groups had worsened compared to their 1995 results.

Analyses of 1997 data (Gamache, Levinson et al, In Press) indicated that the neurocognitive performance of the Eliminators and Forestry workers continued to worsen, and the Forestry workers who resided only 25-39 kilometers from ground zero also began to appear globally impaired. The performance of the Agricultural workers also declined over the first 3 years of the study, although not nearly as much as that of the other 2 exposed groups. Probability levels for all statistical analyses were most frequently at the .001 level of significance.

The analyses of the 1998 data (Gamache, Levinson et al, In Press) indicated that the performance of the 3 exposure groups was leveling off on some of the measures, but continued to decline on others. The performance of the agricultural workers was asymptotic on most measures. Statistically significant correlations between levels of radiation dosage and cognitive performance averaged over the 4 years of the study were observed on 18 of 20 ANAM test scores for the combined exposed groups. The overall results suggested the existence of clinically meaningful neurotoxic effects of ionizing radiation, with the magnitude of the effects being significantly correlated with amount of exposure.

Toluene

Rahill and collaborators (Rahill, Weiss et al. 1966) used ANAM to study the effects of toluene on cognitive performance. These investigators also used SynWork, a multi-task developed by Elsmore (1994) that can be administered within the ANAM test system. In their study Rahill and colleagues exposed 6 healthy subjects to either 100ppm toluene, or air for the control condition, for 6 hours in a double blind randomized fashion. Exposures included 30 minutes of exercise at a level that quadrupled minute ventilation. Exposures were separated by at least 14 days. Toluene was measured in blood and in exhaled air before, during, immediately after exposure, and at 1 and 2 hours post exposure. Lung functioning was measured before exposure and immediately after. Following exercise, the mean blood toluene level was 1.5 μg and the mean toluene level in exhaled air was 28 ppm. During toluene exposure, the composite score from SynWork, which reflects the subject's overall performance on 4 tasks presented simultaneously, was significantly lower than under the control condition ($F = 29.20$, $p = .005$). There was a 10% reduction in this score during the final hour of toluene exposure. With respect to ANAM measures, mean response time increased under the toluene condition for Sternberg Memory Search, Running Memory, Procedural Response Time, and Spatial Processing. Decrements secondary to toluene exposure were also noted for the Sternberg throughput score.

Depleted Uranium

McDiarmid et al. (2000) used the ANAM to examine the effects of depleted uranium (DU). The military employed DU in projectiles and in armor due to its high-

density, availability, and the relatively low cost. DU is a byproduct of the uranium enrichment process. It is approximately 60% as radioactive as natural uranium.

Soldiers wounded in DU-related friendly fire incidents potentially suffered exposure as a result of inhalation or directly by being wounded with this material. During the Operation Desert Storm, 15 Bradley fighting vehicles and nine Abrams tanks were mistakenly fired upon and struck by munitions containing DU. DU is a byproduct of the uranium enrichment process. It is approximately 60% as radioactive as natural uranium a (McDiarmid et al., 2000).

The health effects of concern related only partially to DU's relatively low intensity radiologic toxicity (McDiarmid, et al., 2001). This toxicity is caused by high-energy, but poorly penetrating Alpha emissions. Internalized DU delivers irradiation doses to tissue in immediate contact with it. This effect is a function of contact time, particle solubility, and rate of elimination. There has been no convincing evidence for an increased risk of lung cancer in miners exposed to uranium (McDiarmid et al, 2001). While bone is known to be an important long-term storage depot for uranium, no increased bone cancer rate has been reported in uranium-exposed cohorts (McDiarmid et al 2001). Based on these previous findings of a lack of association between uranium and increased cancer risk and the fact at DU is a heavy metal, uranium's chemical toxicity seemed to be the principal concern for war veterans. As with other heavy metals, there is a particular focus on the kidney along with the respiratory system (McDiarmid et al., 2000).

Some crewmembers surviving these friendly fire instances were left with shrapnel in muscle and soft tissue. These survivors have been followed medically at the VA Medical Center in Baltimore to assess potential untoward health effects. To date, 71 veterans participating in this health-monitoring program have been assessed from one to five times at the Baltimore VA.

Neurocognitive assessment has been part of this general health risk assessment. ANAM has been employed along with traditional test measures as a means of monitoring potential cognitive changes related to exposure to DU. Neurocognitive data have been analyzed in two ways. Veterans exposed to DU were divided into high and low exposure groups based on the amount of uranium found in their urine. This division was based on exposure levels reported in previous research. The drawback to this method is that it is not known at what levels uranium could become toxic. Researchers also attempted to treat the concentration of uranium in the urine as a continuous variable to assess possible relationships between cognitive functioning and uranium body burden. This second method has also been problematic due to the non-normal distribution of uranium values in the cohort under investigation. When veterans were divided into high and low exposure groups, there was no statistically significant difference in their performance either on traditional or computerized measures. No relationship was found between traditional neurocognitive measures and urine uranium values using regression techniques and controlling for psychiatric distress and estimated premorbid intellectual functioning. A statistical trend was found using data from ANAM, but the meaningfulness of this trend requires further exploration. The current plan at the

Baltimore VA is to continue following DU exposed veterans to monitor them for health problems potentially related to DU exposure including the longitudinal analysis of neurocognitive data to assess potential changes in functioning over time.

Cognitive effects at high altitude

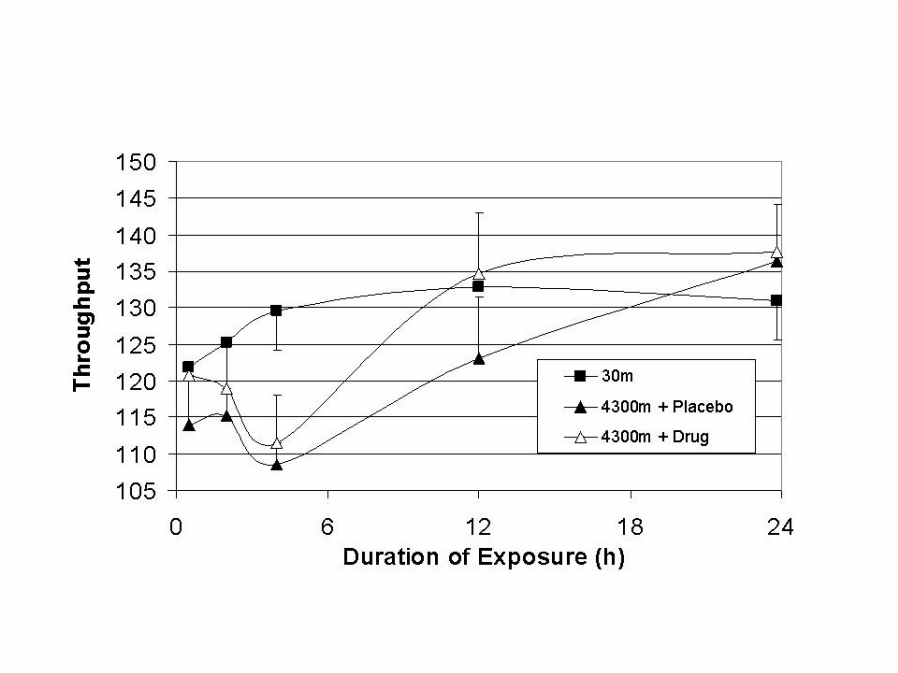
In civilian and military applications, the time course of a disorder greatly influences the impact and costs associated with that disorder. The time course may also suggest the disorder's mechanism or ways to treat the disorder. Banderet, Kane, and Muza (2002) investigated subjective symptoms and cognitive test performance associated with high altitude conditions. These investigators hypothesized that: 1) the time courses of cognitive performance decrements and symptoms of acute mountain sickness (AMS) would be different; i.e., cognitive performance impairments would occur soon after ascent to altitude and be greatly improved after 12 h.; and 2) adverse changes in cognitive performance and symptoms would be decreased by the drug treatment, montelukast. This effort was part of a larger double-blind, cross-over study to assess the role of leukotrienes in high altitude sickness (Muza, Kaminsky, et al, 2004; Muza, Kaminsky, et al. 2001). The investigators used a subset of ANAM measures employed by NASA for use in space flight called the Spaceflight Cognitive Assessment Tool for Windows (WinSCAT) (Kane, Short, et al., in press). The WinSCAT battery was designed to assess the neurocognitive status of astronauts especially during long duration missions. It is currently being used aboard the International Space Station. The WinSCAT battery of ANAM tests includes Code Substitution learning and delayed memory, Running Memory, Mathematical Processing, and Delayed Matching to Sample. The investigators

used WinSCAT throughput scores as their measure of cognitive performance. They also used a computerized 67-item version of the Environmental Symptoms Questionnaire (9 factors) to measure symptoms of AMS.

Cognitive test performance and symptoms of AMS were assessed during the two noncontiguous 48 h periods (24 h at 30 m and then 24 h at 4300 m). Eleven volunteers were treated with montelukast (Singulair™) or placebo in a cross-over design. During each 24 h period, starting at ~0900 h, subjects were assessed after 0.5, 2, 4, 12, and 24 h. Volunteers performed on Panasonic Laptop computers (CF-41 Series); all laptops were powered with 110 VAC. Before the study, the laptops were upgraded to the Windows95™ Operating System and 24 M of RAM, so that they would be compatible with the WinSCAT software.

Cognitive performance impairments were evident on all tasks beginning 0.5 h after ascent and increased through 4 h; at 12 and 24 h, effects did not differ from data collected at 30 m. These trends were most marked for the Code Substitution and the Running Memory Tasks where maximal impairments were 15-20% less than the 30 m baseline values. Impairments on Mathematical Processing and Delayed Matching to Sample were not as marked or distinct as those for Code Substitution and Running Memory. On all WinSCAT tasks, the greatest performance decrements were seen at 4 h. This is illustrated in Figure 1 showing Running Memory throughput performance across the time course for each of the study's test conditions.

Figure 1. Running Memory throughput by time at altitude.



Treatment with montelukast did not improve cognitive performance. The symptoms of AMS were delayed after ascent to high altitude. They occurred after 6 h; they increased in intensity, and were of maximum severity 12 h after ascent. Data using WinSCAT were consistent with those obtained in an earlier study (Banderet & Shukitt-Hale, 2002) in demonstrating the cognitive effects at altitude appeared independent of AMS. In this previous study, the AMS-C factor on the ESQ was not associated with overall cognitive performance,

In this study using WinSCAT, the Running Memory and Code Substitution Tasks appeared more sensitive than the other two performance tasks for assessing cognitive performance decrements after ascent to 4300 m. Decrements in cognitive performance were seen after initial ascent (0.5, 2, and 4 h) and recovered following time at altitude.

Cognitive deficits appeared to resolve by 12 h. Since the authors did not assess subjects between 4 and 12 h, they could not state precisely when test performances returned to normal.

Impairments in cognitive performance had a rapid onset (0-0.5 h) and they recovered before the expression of AMS symptoms. These immediate impairments in cognitive performance with relatively rapid recovery (<12 h) suggest that they are likely mediated by hypoxia. Further studies must be conducted to determine if bold pretreatment strategies can prevent impairments in cognitive performance within the first 4-12 h after ascent to high altitude. If this is possible, some currently prohibited, very advantageous operational applications in US Army flight operations could become possibilities.

Cognitive Assessment in Undersea Medicine

In an effort to adapt ANAM as an assessment tool for U. S. Navy diving studies, a three-year effort was conducted at the Navy Experimental Diving Unit (NEDU). A subset of selected ANAM measures, termed the ANAM Diving Medical Subset (ANAM-DMS), underwent validation studies (Lowe & Reeves, in press) and normative data were established (Lowe & Reeves, 2002). The ANAM-DMS subset included: Subject Demographics, Stanford Sleepiness Scale, Simple Reaction Time, Code Substitution, Code Substitution with both short and long delayed memory, Delayed Matching-to-Sample, Running Memory, and Mathematical Processing. The ANAM-DMS was used in studies of deep saturation diving (Lowe & Reeves, 1999; Lowe & Reeves, 2001),

exposure to warm water diving (Lowe, Reeves & Long, 2002) and accelerated decompression/escape from a disabled submarine (Lowe, & Reeves, 2000).

Saturation Deep Diving

In 1998, NEDU performed a simulated 1000 feet of salt water (fsw) deep dive in their Ocean Simulation Facility (OSF) with the purpose of examining many physiological, medical, and neuropsychological events. For the neuropsychological assessments, Lowe and Reeves (1999, 2000) assessed cognition in nine diver-subjects during a pre-dive baseline, at 500 fsw during compression, twice at the storage depth of 1000 fsw, at the 600 fsw level and the 200 fsw level, and then at the surface post-dive. The greatest effects (impairments) appear to have occurred between the pre-dive and 500 fsw level for accuracy, response time and throughput scores on the Delayed Matching-to-Sample and Code-Substitution long delay memory tasks. Also observed were decreased response time and throughput scores on Code-Substitution immediate memory and a slowing of Simple Reaction Time. Though the decrement still was present at 1000 fsw for each of the measures mentioned above, there appeared to be no decrement greater then that at 500 fsw. The interesting result from this study was demonstrated on the fourth day at 1000 fsw, when the measures from the aforementioned measures all returned to at least baseline levels. This acclimation, like that seen at altitude, continued through each successive assessment concluding with the post-dive measurement.

Data obtained form ANAM measures demonstrated some cognitive changes associated with saturation deep diving; however, the underlying mechanisms remain to be determined. Possible reasons for these changes include the effects of pressure, the gas mixture, workload, the hyperbaric environment, or a combination of these factors. This study also demonstrated the capability to assess an individual in a hyperbaric chamber, thus providing the medical diving community with a method and means of tracking recovery from a diving insult while the individual is receiving a recompression treatment.

Water Temperature

In 1999, NEDU began to look at the effects of warm water on divers, both physiologically and cognitively. A study by Lowe, Reeves, and Long (2002) examined five water temperatures, 78, 92, 94, 96, and 101 Fahrenheit, under a working condition (peddling a bicycle at 80% Vo2max) or a resting condition. Again, the ANAM-DMS was used to assess diver cognitive performance before and after the dive.

There were significant decreases in accuracy scores on all three parts of the Code-Substitution test (Code Substitution ($p = .018$), Code-Substitution with a short delay ($p < .0001$), and Code-Substitution with a long delay ($p < .0001$)) that were independent of both water temperature and working mode. Mean reaction time and throughput remained relatively unchanged, except for a marginally significant improvement in mean reaction time for Code-Substitution with a short delay ($p = .05$). Running memory accuracy and throughput each demonstrated a significant decline

from pre-dive to post-dive ($p = .001$ and $p = .03$ respectively), and again this decrease was independent of water temperature and work mode. Matching-to-Sample throughput showed diving related decrements that were also independent of water temperature and work mode ($p = .015$). However, mean response time for Matching-to-sample increased marginally with diving and demonstrated a significant ($p = .042$) temperature interaction with the greatest increase in response time seen at the 96.5° water temperature. There were no significant changes in Simple Reaction Time or Mathematical processing.

Only water temperature interacted with Matching-to-Sample mean reaction time. Other trends were not noted. However, the statistical power for the tests was low due to small sample sizes and unequal groupings of subjects at each test matrix (temperature and work mode). Resolution of effects could be improved with a larger sample. As mentioned, diving-related cognitive changes are subtle and statistical power is an important factor in their detection. Future studies could benefit from the inclusion of a control group to aid in differentiating the effects of fatigue and water temperature.

Another limitation of this study was that medical termination criteria used to discontinue a diver's participation in the study impacted data collection. The diver would remain in the matrix condition until the maximum time elapsed (8 hour resting mode, 4 hour working mode), or until voluntary termination of the trial. Divers are very competitive and they often would not terminate until physiologically required to do so (they would pass out or core body temperature would reach a critical medical threshold). Some individuals who reached their physiological threshold subjectively

reported dramatic cognitive changes, such as confusion and hallucinations. However, assessment of these divers was not possible until they were medically stable. The recommended solution for this was to conduct periodic in-water assessments so that cognitive functioning could be obtained as the diver's condition began to deteriorate.

Assessment of Cognitive Changes Associated with Accelerated Decompression Procedures

Between 1998 and 2000, NEDU conducted a series of Nitrox saturation dives that simulated the environment on a disabled submarine that had pressurized to a theoretical depth of two to three atmospheres (33 to 66 fsw). These dives specifically explored accelerated decompression procedures for bringing submariners back to surface while safely providing therapeutic decompression. These studies were conducted over three phases, testing various decompression parameters (treatment gas exposure, frequency of air break, air break durations, and schedules based on decompression models).

Currently approved Navy procedures require a gradual decompression from saturation of at least 33 hours. It is desirable to reduce this time to approximately six hours to match the technical capabilities of the current disabled submarine rescue vehicle (DSRV). The human body is considered to be fully saturated after exposure to pressure of 72 hours. Using 100% (>95%) humidified oxygen as a treatment gas, the goal of six to eight hours decompression was successfully achieved. Phase II examined recomputed

decompression schedules based on existing models. The change between Phase I and Phase II resulted in slightly longer exposures to humidified oxygen. ANAM-DMS was used to assess potential cognitive changes associated with the use of these different air mixtures during rapid assent.

The ANAM-DMS was administered at baseline and following each decompression. Lowe and Reeves (2000, in press) and Lowe, Reeves and Kane (2000) reported on two of the three phases. There were a total of 48 Navy divers in phase one and 73 participants in phase two. For Phase I, mean response times for Code Substitution long delay memory and Matching-to-Sample, and the accuracy measure for Mathematical Processing, demonstrated non-significant declines in performance from baseline. Mean response time and throughput for Running Memory ($p = .003$ and $p = .001$), mean response time for Mathematical processing ($p = .004$), and Code Substitution long delay accuracy ($p = .03$) all demonstrated significant decrements for the Phase I group.

During Phase II, all three measures (mean response time, Accuracy and Throughput) for Code Substitution long delay memory, showed non-significant decrements from baseline. The accuracy measure for Mathematical Processing and the mean response time for simple reaction time showed non-significant decrements. Matching-to-Sample yielded the only significant decrements in performance on both the mean response time ($p = .01$) and throughput ($p = .011$).

The ANAM-DMS instrument was sensitive to diving related insults. However, since the ANAM-DMS was being evaluated for possible use as a screening tool, treatment decisions during these studies were based on medical criteria and on the subject presenting with symptoms of decompression sickness (DCS). However, Lowe, Southerland, and Reeves (1999) and Reeves, Lowe, and Levinson (2001) presented a case that reported with cognitive symptoms requiring treatment for DCS. As a member of the saturation dive team for the disabled submarine project, the subject received a baseline and a post-dive assessment. An increase in performance scores was noted on all traditional test measures post-dive. This finding is similar to pre- and post-dive comparisons of group performance where all measures had significant ($p \leq .004$) increases in performance except for Trail Making Form B, which showed trend ($p = .06$) toward improved performance. During the post-dive medical physical examination, no symptoms were reported. The next day, the diver reported symptoms including a decrease in peripheral vision, lack of sustained concentration and a general mental slowness. The results of the neurological exam by the attending physician were within normal limits. Based upon the presentation of the symptoms, a U.S. Navy treatment table six was completed with full resolution. A review of the post-dive ANAM-DMS results showed statistically significant performance decrements from baseline on matching-to-sample as evidenced by an increase in mean response time (RT) (1790.053 ms to 2396.113 ms) and a decrease in mean throughput (29.76 to 23.60). By comparison, the peer group had a non-significant increase in mean RT from 1740.810 ms to 1792.992 ms and a non-significant decrease in throughput from 33.288 to 32.841. Two days post-treatment the diver completed a follow-up assessment

that showed performance on Matching to Sample better than baseline; mean RT decreased to 1701.023 ms and the mean throughput increased to 35.62

Although it would have been desirable to get cognitive assessment data immediately before treatment of a symptomatic diver, this was not possible due to time constraints and the urgency to treat DCS upon presentation of any symptoms. However, this case was of interest since Matching to Sample appeared sensitive to dive effects in the above studies and in this case may have been prodromal to the onset of more apparent DCS. A second case, similar to this, occurred later. It differed in that the diver's symptoms were noted sooner and he demonstrated performance decrements on other ANAM measures. Additional data are needed to establish if such changes in ANAM performance might be useful in the early detection of DCS.

Cognitive changes during Marine basic training

Basic training is inherently stressful. Fatigue, physical and mental stress are the primary issues. Gastaldo, Reeves, Levinson and Wenger (2001) studied Marine Corps recruits during basic training and established norms for this "young adult" population. These investigators employed a customized subset of an earlier MS DOS version of ANAM to assess 197 Marine Corp recruits. This ANAM subset consisted of the Stanford Sleepiness Scale, the Walter Reed Mood Scale-2, Code Substitution with immediate and delayed memory, Sternberg Memory Search, and two versions (80 and 160 items) of the Running Memory Continuous Performance Task, Digit Set Comparison, Logical Reasoning, Matching to Sample, Mathematical Processing, Spatial Processing, and Simple Reaction Time Tasks repeated up to 3 times. Approximately one month later,

retest data were collected on 50 of these individuals. Normative data for females (n = 83) and males (n = 114) and retest data were obtained from this study. .

Gastaldo, Reeves et al. (2001) also followed 14 recruits who displayed symptoms associated with delusional hyponatremia during training. The Marine Corps normative data were used to assess the effects of this malady on the group as a whole and to provide a clinical basis for evaluating each case. For the 14 hyponatremia cases, a one sample t-test was used to assess differences between these cases and the normative group. During the initial assessment, which occurred within 24 hours of the onset of heat stress symptoms, mean throughput scores on 8 tests and mean accuracy scores on 3 tests were significantly lower for the 14 cases as compared to controls. These scores returned to normal at follow-up, which occurred 1 to 4 weeks post treatment. Gastaldo, Reeves et al. (2001) noted that this pattern of performance is similar to that exhibited by the traumatic brain injury (TBI) patients presented by Levinson and Reeves (1996). However, unlike the TBI patients, the hyponatremia cases returned to normal levels at retest.

Military personnel cognitive changes during high stress exercises

A series of studies was conducted to examine the effects of operational stress on military personnel cognitive performance. In the first study, Harris, Hancock & Harris (in press) examined the cognitive effects of an extreme training environment, the Navy Survival, Evasion, Resistance, and Escape (SERE) School at Brunswick, ME. Cognitive performance of thirty-five Navy and Marine personnel was assessed with the ANAM before and after seven days of intense field training. Simple Response Time

increased 10.4 ms after training (p <.05), but other cognitive tasks were stable following completion of the experience. However, a more detailed analysis of the data indicated an apparent inability to maintain performance as time-on-task increased. Spatial Processing and Logical Reasoning performance on items near the end of scales was less accurate (p < .001, and .01, respectively) and Code Substitution, Logical Reasoning, and the Continuous Performance Task response time became longer as the end of each test approached (p < .001, .01 and .01, respectively). Post-study debriefing suggested that a significant degree of recovery may have occurred during the three hours between the end of training and assessment, raising the question of whether larger changes may have been detectable during training. Results suggested that increased Simple Response Time and the inability to maintain performance should be further evaluated as potential early indicators of impairment in stressful situations.

Studies two and three examined cognitive performance during simulated infantry combat exercises (Harris & Hancock, 2003; Harris, 2004). A subset of ANAM tests was administered to soldiers before the exercises began to establish a pre-stress baseline and periodically during the simulated combat exercise to examine the effects of training. Test measures were administered on a PDA (ANAM Readiness Evaluation System, ARES).

In the first infantry study, cognitive performance did not change during a five day exercise in which the soldiers' main task was to delay the advance of an invading force. As assessments were conducted by the experimenter, it was possible to observe

operating conditions of each soldier tested. Although two companies were participating in the same exercise, it was found that the environmental and workload conditions varied. One Company operated along the first section of a road in an area with heavy underbrush and steep, rocky ridges, and was at the point of attack. Continuous movement by these soldiers was required during a series of engagements, and because no breeze penetrated the woods, there was little relief from the 90 plus degree heat. The other Company defended a pass further along the road, which the invading force never reached. This group did not move, and with little underbrush on their ridge, they benefited from any breeze that occurred. Comparison of test scores in soldiers in the high and low demand settings suggested that stress level and conditions may have affected cognitive performance changes, but sample sizes were insufficient to adequately assess this hypothesis.

Soldiers returned to the barracks area for three days before initiating an assault on week two, day four. The only cognitive performance change noted following this 27 hour assault mission was a 9.2 ms increase in Simple Response time compared to baseline ($p < .05$).

During an infantry exercise the following year, the environmental conditions and mission demands of all participating soldiers was similar. The terrain was heavily wooded and all companies were attacked simultaneously. In-field assessment during the defensive operation on exercise Day 3 indicated that soldiers' Simple Response Time had increased 13.4 ms from their pre-exercise level ($p < .01$) and response time on three

information processing tasks, Logical Reasoning, Spatial Processing, and the 6-item Sternberg task, increased ($p < .01$, .001, and .01, respectively). There was no indication of decreased accuracy on Day 3. Information processing response time returned to baseline levels at the end of the five-day defensive exercise. Simple Response Time remained elevated, 9.4 ms above baseline, but the difference was not statistically significance ($p < .07$). However, accuracy on two tasks, Logical Reasoning and Mathematical Processing, were below baseline levels on Day 5 ($p < .05$ and .05, respectively).

Following their defensive operations, companies remained in the field for one day while preparing for an operation in which they would assault the opposing force's position. After the 27 hour assault operation, Simple Response Time was approximately 21 milliseconds above baseline ($p <.05$), approximately twice the 9.2 ms. post-assault change observed in the first infantry study. Logical Reasoning and Mathematical Processing response time after the operation were also greater than baseline level ($p < .05$ and .05, respectively).

Assessing cognitive performance during field operations is a challenging task and studies need to account for the fact that stress levels and environmental challenges will vary depending on specific conditions and assignments. Taken together, data from these studies suggested that changes in response time may be an early indicator of cognitive change. Changes during field operations appeared to reflect operational conditions with increased response times following periods of high operational

demands. Complex cognitive task performance did not appear to be affected in the above studies until Simple Response Time increases approached 15 ms above baseline. Increased errors occurred during a low demand period following a five day defensive operation.

Fatigue and Cognition

As noted above, WinSCAT is a special implementation of ANAM designed to assess the cognitive health of astronauts. In a NASA sponsored study, French, et al. (1999) evaluated the sensitivity of WinSCAT to one of the pre-dominant stresses facing astronauts, fatigue induced by sleep deprivation and circadian disruption. Two groups of eight US military pilots (ages 30-40) participated in this study of the performance effects of 46 hrs of sleep deprivation -- a period that extended over two circadian performance nadirs. In addition to WinSCAT, four cognitive performance tests typically used in fatigue research were performed repetitively during the sleep deprivation period. WinSCAT was performed every six hours up to the 36th hour of sleep deprivation.

In their initial analysis, French et al. (1999) analyzed only the 6 scores from the WinSCAT battery that currently appear as immediate feedback to the astronauts. The feebback includes both accuracy and reaction time scores. For all tests, sleep deprivation's greatest effect was on response time. With respect to the WinSCAT tests, Mathematical Processing exhibited significant fatigue-related decrements on the last 3 of the 6 trials, following 23 hrs of wakefulness. Running Memory and Matching to Sample showed significant decrements on 2 of the trials, after 23 and 29 hours awake.

Of all the tests, only Continuous Processing demonstrated an effect on response accuracy, and the effect was associated with 23 and 29 hours of wakefulness. In contrast, two of the comparison tests demonstrated sensitivity on 21% of the trials and another on only 17% of the trials. WinSCAT was found to be sensitive to fatigue early (within 23 hours) and late (35 hours) into the sleep deprivation period. It is likely that WinSCAT will identify astronauts too fatigued to optimally perform sensitive missions.

Since the analysis of French and his colleagues was limited to selected accuracy and response time scores, Kane, Short et al. (in press) reanalyzed these data using throughput scores. With respect to throughput scores, significant decrements in performance were noted at 23, 29, and 35 hours of sleep deprivation for Running Memory and Delayed Matching to Sample; and at 29 and 35 hours for Mathematical Processing.

Kane, Short et al. (in press) also analyzed individual change scores using the Reliable Change Index method (Jacobson and Truax, 1991) to look for individual differences in fatigue responses among the study participants. For this study, participants took WinSCAT 10 times to control for practice and to obtain a baseline before entering the study conditions of continued wakefulness. However, operationally, WinSCAT is given 6 times before flight with the final 3 administrations serving as baseline. Consequently, Kane and Short used throughput scores obtained during trials 4 through 6 to compute the RCI then tabulated the number of scores for each participant that fell outside of the expected range at each test interval.

Table 1 presents data obtained from 11 to 42 hours of wakefulness. Values in the table represent the number of scores falling in the impaired range based on the RCI at each test interval. Four of the 5 WinSCAT scores were included. Data from the Code Substitution Delayed Recognition Test are still being analyzed. Participants fell into three groups. The first group performed well throughout the entire study and had from zero to one score impaired despite being awake an extended time without sleep. Group two evidenced periods of marginal performance with two scores in the impaired range at different time points. Group 3 demonstrated the most severe changes in performance having 3 or 4 tests were below expectations based on the RCI. Not all subjects completed all trials.

Table 1 Number of impaired scores by participant at various hours of sleep deprivation

Group	11hrs	17hrs	23hrs	29hrs	35hrs	42hrs
1	0	0	0	0	1	0
1	0	0	0	0	0	0
1	0	0	1	0	0	
1	0	1	0	1	0	0
2	0	0	1	2	1	1
2	0	0	1	2	2	0
2	2	1	2	2	1	
3	1	3	3	3	3	
3	1	2	4	4	2	
3	0	0	3	3	1	1
3	1	0	0	4	2	0
3	0	2	4	4	1	4

The above study assessed fatigue brought about by sleep deprivation; whereas, Law and collaborators (2003) assessed fatigue brought about by physical exercise. These investigators studied 20 healthy moderately to highly physically fit males

(ages 22-40, education 15-20 years). All participants took repeated trials of WinSCAT to stabilize baseline performance. During the experimental portion of the study, WinSCAT was administered before and after a load carriage (30% bodyweight backpack) exercise endurance (70% VO_2 Max x 2hrs) treadmill test. Participants were given two trials (Tyrosine and placebo experimental trial conditions) separated by at least a week. Four subjects produced invalid data on the experimental study trials (administration/equipment errors), leaving a final study sample size of n=16 (ages 20-45, education 16-20 years). DEX suppression classification of High or Low stress responders was based on degree of change in adrenocorticotrophin hormone (ACTH) levels collected before and immediately after the treadmill stress test, approximately 8 hours after DEX ingestion. WinSCAT scores showed deterioration due to exercise fatigue, however, no differences were found between subjects based on there DEX response or when attempting to use Tyrosine as a countermeasure.

CONCLUSIONS

The necessity and benefits of exploring harsh environments are very real. It is sometimes necessary to send humans into space, deep-sea environments, extremely cold or hot environments, and settings that are contaminated with hazardous materials, and it is often necessary that they perform complex cognitive functions while there. Knowledge of the immediate effects of these conditions on cognitive functioning and the long term neuropsychological consequences of exposure are needed to intelligently

manage personnel in extreme environments. In military settings, the warfighter is asked to perform cognitively demanding tasks during long, sustained missions. The operation tempo in Iraq and the resulting fatigue illustrate this problem.

In this paper we reviewed studies using ANAM to assess changes in cognition associated with different stressful conditions. The purpose was to bring together and summarize work for diverse stressors and environmental conditons. In general, ANAM measures appeared sensitive to effects of various conditions and diverse environmental stressors. These findings are not surprising since ANAM is an outgrowth of efforts within DoD to develop automated test batteries for assessing human performance. Nevertheless, future studies using ANAM are needed for its validation.

There are two basic questions of military relevance for all studies of environmental stressors. The first is to characterize the expected effects of a stressor on a group of individuals. The second is to develop methods of identifying when a given individual is showing appreciable changes in performance and cognition. Most ANAM research efforts were classical studies looking for group effects in response to an environmental stressor. It is well understood, however, that there is significant variability in how individuals respond to stressors. Work has begun within the ANAM community to study individual change measures and to explore statistical techniques for identifying when a specific individual develops cognitive decrements that are a cause of concern. The saliency of this effort is evident from the individual case reports noted in this review by Lowe and his colleagues where the possibility existed that changes in cognition were

prodromal of DCS. Addressing the second question will likely require approaches that combine collecting reference group norms with statistical methods of assessing significant change for an individual. Substantial work remains to be done in assessing individual change. The reader is referred to the paper in this special issue by Roebuck-Spenser and colleagues which discusses change score issues.

With respect to ANAM, data are now available pertaining to diverse stressors including atmosphere pressure, various aspects of fatigue, psychological stress, and toxic insults. Data support the use of ANAM to detect deterioration in cognitive performance and to monitor recovery. Other papers in this issue address the use of ANAM in the study of cognitive changes secondary to disease and injury. Hence, current data appear to support the use of ANAM as an assessment tool contributing to the medical blueprint of the warfighter. This information provides a characterization of cognitive performance in extreme environments so that human performance guidelines can be established for operating in hostile environments.

REFERENCES

Banderet, L.E., Kane, R.L., & Muza, S.R. (2002). *Time course of cognitive performance after ascent to 4300 m.* Presented at APA's Annual Midyear Symposium, Division 19 (Military Psychology).

Banderet, L.E., & Shukitt-Hale, B. (2002). Cognitive performance, mood, and neurological status at high terrestrial elevation. In *Textbook of Military Medicine: Medical Aspects of Harsh Environments*. Washington, DC: Borden Institute, 2, 729-763.

Elsmore, T. (1994). "SYNWORK 1: A pc-based tool for assessment of performance in a simulated work environment." *Behavioral Research Methods, Instruments, and Computers, 26*, 421-426

Flynn, C., R. Kane, et al. (2000). *Background and Development of the Space Flight Cognitive Assessment Tool*. 71st Scientific Meeting of the Aerospace Medical Association, Houston.

Gamache, G.L., Levinson, D.M., Reeves, D.L., Bidyuk, P.I., and Brantley, K.K. (2005). Longitudinal neurocognitive assessments of Ukrainians exposed to ionizing radiation after the Chernobyl nuclear accident. *Archives of Clinical Neuropsychology, 20* (1), 81-93

Gastaldo, E., Reeves, D., Levinson, D., & Wenger, B. (2001). ANAM USMC normative data: effects of heat stress on neurocognitive performance. *Archives of Clinical Neuropsychology, 16* (8), 725.

French, J., Neville, K. J., Eddy, D., Storm, W., & Flynn, C. (1999). *Sensitivity of SCAT to Sleep Deprivation: NASA Technical Report*. Limited Distribution

Harris, W.C. (2004). *Stress effects on soldier performance*. Proceedings of the Human Factors and Ergonomics Society 48th annual meeting, New Orleans, LA.

Harris, W.C., Hancock, P.A.& Harris, S.C. (In Press) Information processing changes following extended stress. *Military Psychology*.

Harris, W.C. & Hancock, P.A. (2003). *Field assessment of cognitive performance under stress*. Proceedings of the Human Factors and Ergonomics Society 47th annual meeting, Denver, CO.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol, 59*(1), 12-19.

Kane, R.L., Short, P. Sipes, W., & Flynn, C.F. (In Press). Development and validation of the Space-flight Cognitive Assessment Tool for Windows (WinSCAT). *Aerospace and Environmental Medicine*.

Law, W.A., Kane, R.L., Ashburn, J.R., Kim, S.J., & Deuster, P.A. (2003). Cognitive inefficiency after exercise-stress: Resistance to tyrosine enhancement. *J International Neuropsychological Society*.

Levinson, D.M. & Reeves, D.L. (1996). Monitoring recovery from traumatic brain injury using the Automated Neuropsychological Assessment Metrics. *Archives of Clinical Neuropsychology, 11*(5), 419.

Lowe, M.A., Reeves, D. L., & Kane, R. (In Press) *Construct validity of the automated Neuropsychological Assessment Metrics (ANAM) Battery used in U. S. Navy Diving Research.* Naval Hospital Technical Report. Camp Pendleton, CA.

Lowe, M. A. & Reeves, D. L. (In press). *Accelerated Decompression From Nitrox Saturation Using 100% Oxygen: Neuropsychological Findings. NEDU Technical Report*, Panama City, FL: Navy Experimental Diving Unit

Lowe, M.A., & Reeves, D.L. (2002) *Automated Neuropsychological Assessment Metrics: Norms for U. S. Navy Divers, NEDU TR 02-05*, Panama City, FL: Navy Experimental Diving Unit.

Lowe, M.A., Reeves, D.L., & Long, E. (2002) *Warm Water Diving 1999: Neuropsychological Findings*, NEDU-TR-02-11, Panama City, FL: Navy Experimental Diving Unit.

Lowe, M. A., & Reeves, D. L. (2001). *Deep Dive 1998: Neuropsychological Findings. NEDU TR 13-01*. Panama City, FL: Navy Experimental Diving Unit.

Lowe, M. A., & Reeves, D. L. (2000). "Accelerated Decompression From Nitrox Saturation Diving: Neuropsychological Findings." *Undersea and Hyperbaric Medicine*, Vol. 26 Supplemental, p. 18.

Lowe, M. A., & Reeves, D. L. (1999). "At depth evaluation of a computerized neuro-pyschological assessment tool: Deep dive 98." *Undersea and Hyperbaric Medicine*, Vol. 26 Supplemental, p. 18.

Lowe, M. A., Reeves, D. L., and Kane, R. (2000). "At depth computerized assessment of neurocognitive changes in divers." *Archives of Clinical Neuropsychology, 15*(8), p. 699.

Lowe, M.A., Southerland, D. G., & Reeves, D. (1999). "Computer-Based neuropsychological findings in a patient with dcs following accelerated decompression from a saturation dive: A case study." *Undersea and Hyperbaric Medicine, 26* Supplemental, p. 17.

McDiarmid, M.A., Keogh, J.P., Hooper, F.J., McPhaul, K. Squibb, K., Kane, R., DiPino, R., Kabat, M., Kaup, B., Anderson, L., Hoover, D., Brown, L., Hamilton, M., Jacobson-Kram, D., Burrows, B., Walsh, M. (2000). Health effects of depleted uranium on exposed Gulf War Veterans. *Environmental Research, section A*, 82, 168-180.

Muza, S.R., Kaminsky, D., Fulco, C.S., Banderet, L.E., and Cymerman, A. (2004). Cysteinyl leukotriene blockade does not prevent acute mountain sickness. *Aviation, Space, and Environmental Medicine, 75*, 413-419.

Muza, S.R., Kaminsky, D., Fulco, C.S., Sonna, L.A., Banderet, L.E., Kane, R., Cymerman, A. (2001). *Effect of cysteinyl leukotriene blockade on development of acute mountain sickness*. Technical Report, T-01/12, US Army Research Institute of Environmental Medicine, Natick, MA.

Rahill, A., B. Weiss, et al. (1966). "Human performance during exposure to toluene." *Aviation, Space, and Environmental Medicine, 67*(7): 640-647.

Reeves, D., Lowe, M., & Levinson, D. (2001). Comparative results of traditional and computerized tests in a case of decompression sickness following a saturation dive. *Archives of Clinical Neuropsychology, 16*(8), 733.

COGNITIVE CHANGE IN SPECIAL FORCES PERSONNEL FOLLOWING STRESSFUL SURVIVAL TRAINING

W. C. Harris, P. A. Hancock
University of Central Florida

C. A. Morgan III
Yale University School of Medicine

Understanding the deterioration in cognitive functioning produced by stress continues to gain in importance due to the increasing demands imposed by technologically sophisticated systems. Although the general deleterious effects of stress are well established, the relative sensitivity of different cognitive functions to stress and the pattern of cognitive recovery with rest have not been fully distinguished. In this paper, we examined the cognitive performance of Special Forces soldiers immediately prior to and immediately following one week of Survival, Evasion, Resistance and Escape (SERE) School training at Ft. Bragg, NC. Post-stress cognitive performance was characterized by significantly increased response time with minimal change in response accuracy. While response time increased for all tasks, memory appears to be most sensitive to stress. Performance returned to pre-stress levels the next morning following one night of sleep. The tasks affected most in the current study differed from changes which follow primarily upon physical stress, implying that the effects of combined psychological and physical stress on cognitive performance differ substantively from the effects of physical demand alone.

INTRODUCTION

Modern day military forces are currently operating in a variety of stressful settings for extended periods of time (Scales, 2003). These chronic stress conditions can impair cognitive functioning which is the basis for the judgments these personnel take. Errors in such judgments can be personally and collectively catastrophic. As decision-making ability declines, we need to ensure that critical tasks can still be performed and that technological systems adjust to support this reduced level of operator functioning (Hancock & Szalma, 2003). Predicting performance change during stress requires our understanding of the function relating change in cognitive capacities to stress. The extended-U model (Hancock & Warm, 1989) provides a comprehensive description of this complex, non-linear relationship between stress level and operator performance. However, this general description does not specify the nature of change in any individual cognitive function as stress increases. Judgments, and the decisions which underlie those judgments, are the final product of a number of such basic cognitive functions (Wickens & Flach, 1988). The cognitive operations involved in decision-making include; i) obtaining relevant information, and/or retrieving relevant information from memory; ii) processing logical and spatial information as well as mathematical transformations; and iii) formulating a chain of appropriate response strategies. At present, we do not know whether stress affects the resources necessary to perform each of these functions equally or whether they are differentially altered by either stress or the inherent characteristics of the exposed individuals. Predicting impending overall performance deterioration is of course important. However, if we are to provide technical support to such stressed individuals, we need to be able to anticipate which cognitive functions will deteriorate most.

Cognitive changes during high stress portions of SERE training include dissociation and memory deficits (Morgan, et al., 2004). Previous field assessments, which have examined only a limited number of cognitive tasks following approximately 12 hours of high stress and a recovery period of approximately one hour, have detected little in terms of specific cognitive changes (DeJohn & Reams, 1992). However, the subsequent inclusion of Simple Response Time (SRT) and a more detailed analysis of within-session did find subtle cognitive changes following one week of high stress and only three hours of recovery (see Harris, Hancock, & Harris, 2005). Further, in some of our previous work, we have also found small changes in cognitive performance immediately following physically demanding work (Harris & Hancock, 2003). The present study examined the effect of high levels of combined physical and psychological stress immediately following SERE exposure and also monitored the recovery profile as it occurred during a twelve-hour, post-stress interval. The cognitive battery selected provides a profile of six cognitive functions critical for decision making in military settings. Previous studies (Harris et al., 2005; Harris & Hancock, 2003) found that SRT changes preceded response time change in more complex tasks and the number of milliseconds by which SRT increased, actually predicted the level of deterioration of the more complex tasks. The present study served to further evaluate this covariation of SRT-cognitive functioning.

EXPERIMENTAL METHOD

We assessed the performance of two classes at the Army SERE School at Ft. Bragg, NC using the ANAM Readiness Evaluation System (where ANAM is the acronym for the Automated Neuropsychological Assessment Metric) to evaluate soldier cognitive functioning. The major goal of the present experiment was to determine whether all types of cognitive functioning are affected equally by a common level of mission stress, namely one week of SERE training. We also examined the recovery of cognitive functioning after this period of stress. The results of each class were analyzed separately and compared.

Experimental Participants

Two SERE classes were solicited as participants in the present study. During December 2003, twenty-seven enlisted Special Forces soldiers enrolled in the Army Survival, Evasion, Resistance, and Escape (SERE) School at Ft. Bragg, NC, and represented the first class. In January 2004, a second group of twenty-four Special Forces soldiers enrolled in the Army SERE School and participated as the second class. Twenty-two of the soldiers in the second group were enlisted personnel and two were officers.

Cognitive Assessment Methodology

The battery used to assess performance in the present study was the ANAM Readiness Evaluation System (ARES). The Automated Neuropsychological Assessment Metric (Reeves, et al., 1991) presents a wider spectrum of cognitive tasks on a desktop or laptop computer, making administration in the field problematic. The ARES is a selection of tests from the larger ANAM battery presented on a handheld computer. ARES tasks were selected that assess components of cognitive performance considered to be important in operational settings (Wickens & Flach, 1988). The battery included; Simple Reaction Time (SRT), Logical Reasoning, Mathematical Processing, Spatial Processing, Spatial Memory, and a 6-item Sternberg Memory task. The battery was always administered using a handheld, Handspring Visor Prism.

Experimental Procedure

Training on the cognitive assessment battery (ARES) and subsequent baseline assessment occurred at approximately 1800h on the last day of SERE classroom instruction. During this pre-stress session, soldiers completed the battery four times. Collectively they exhibited high levels of accuracy during the first training trial and accuracy was maintained while response speed increased during training. Response time improved with practice and the fourth repetition of the respective ARES tasks was used as baseline performance (see Benedetto, Harris, & Goernert, 2000). One week later, at approximately 1400h, soldiers completed the battery a fifth time as their final activity of the field training exercise. Meals were provided and snacks were available, and soldiers then slept in the barracks that night. The sixth and final ARES administration occurred at approximately 0715h the following morning.

EXPERIMENTAL RESULTS

Although procedures were identical for both classes, results are presented separately to provide an estimate of the consistency of effects. We have divided the results initially on the basis of response accuracy and response time.

Response Accuracy

Pairwise comparisons were conducted between the fourth performance trial (designated as the baseline), the fifth performance trial (designated as the immediate post-stress condition) and the sixth performance trial (designated the recovery, or morning after trial). For Class 1, there was only one significant effect and that was the reduction in accuracy of Logical Reasoning for the morning after trial versus the baseline, see Table 1. These predominantly null results for accuracy follow our earlier finding of relative high stress resistance in the real-world. The results of Class 2 were a little different. Here, the accuracy of mathematical processing was impaired in both the immediate post-stress and morning after test conditions. Further, there was a transient effect on Letter Memory which showed a small, but significant drop immediately following stress. However, like the results for Class 1, the predominant pattern showed little or no substantive change in performance accuracy.

Table 1. *Accuracy by Task and Condition.*

Task	Class	Pre-stress	Post-stress	Morning after
Logical Reasoning	One	94.6	94.2	91.2 *
	Two	95.0	94.3	94.1
Mathematical Processing	One	91.0	90.5	92.5
	Two	94.5	88.2 *	89.8 *
Spatial Processing	One	95.0	91.2	91.7
	Two	95.8	92.3	92.6
Spatial Memory	One	94.7	92.2	94.7
	Two	96.0	93.9	96.7
Letter Memory	One	94.7	92.2	94.7
	Two	96.7	93.6*	97.9

*p< .05

Response Time

While there appear to be few changes in response accuracy one cannot take this as a full portrait of performance since speed is symbiotic with accuracy and one cannot be veridically understood without the other. When we look at the results in Table 2, a very different and much fuller picture emerges. The first thing that must be mentioned is that the present results do not derive from a speed-accuracy trade-off. It is evident that when performance was less accurate it was also slower, implying a general deterioration, not a change in response strategy. For Class 1, there was a significant post-stress increase in response time for all tasks, with the exception of Spatial Processing which however did show a strong tendency for slower response. This post-stress effect was transient in that by the morning after, scores had all returned to baseline. For Class 2, the pattern was almost identical. However, in this group the exception was Logical Reasoning and not Spatial Processing. Again the trend for this one non-significant change was in the same direction as all of the others, namely an increase in response time. Combinatorial examination of both the time and accuracy data did not suggest any micro speed-accuracy trade-offs within any of the individual tasks tested.

Table 2. *Response Time by Task and Condition.*

Task	Class	Pre-stress	Post-stress	Morning After
Simple Response Time	One	215.4	236.1***	213.2
	Two	224.8	247.1*	224.0
Logical Reasoning	One	1761.7	1898.3*	1801.1
	Two	2106.4	2224.7	1915.8
Mathematical Processing	One	1866.8	2234.3***	1829.3
	Two	2124.8	2590.8***	2306.6
Spatial Processing	One	2073.8	2284.4	2000.0
	Two	2407.0	2997.5*	2239.0
Spatial Memory	One	1168.9	1692.0***	1180.0
	Two	1287.8	1681.6**	1429.4
Letter Memory	One	690.0	845.5***	721.2
	Two	778.8	1039.9***	780.9

p< .05, ** p< .01, *** p< .001

Cognitive Function Sensitivity

One of our prime concerns was to estimate the magnitude of stress effects on specific information processing elements. To generate a first pass perspective on this, we reasoned that Simple Response Time (SRT) was involved in all tasks and so we have extracted SRT changes from all other members of the task battery. (This assumes a simple additive effect of SRT which we are willing to make at this juncture). This being so, the response time changes are presented in Table 3. Similar changes in both classes on four of the five tasks suggest that the effect of stress varies with largely the task and is somewhat loss contingent upon the individual. The size of post-stress change varied with the cognitive demands of the task. Memory tasks were affected most. For Spatial Memory this change was 53% and 35% for Class 1 and Class 2 while the respective reduction in Letter Memory was 28% and 43% respectively. The Logic Reasoning task was affected least (7% and 5%). Mathematical Processing also exhibited similar, but more moderate changes in both classes (21% and 23%). The only task not exhibiting similar changes in both classes was Spatial Processing (i.e., 10% and 26%). Performance on the Spatial Processing task stabilizes slowly (Harris et al., 2005) and the class difference may reflect that performance had not stabilized during training in one or both classes.

Table 3. *Response Time Change by Task and Condition.*

Task	Class	Pre-stress	Post-stress	Change (ms)	Change (%)
Logical Reasoning	One	1547	1662	115	7
	Two	1881	1978	97	5
Mathematical Processing	One	1652	1998	346	21
	Two	1900	2344	444	23
Spatial Processing	One	1859	2048	189	10
	Two	2182	2751	569	26
Spatial Memory	One	954	1456	502	53
	Two	1063	1435	372	35
Letter Memory	One	475	610	135	28
	Two	554	793	239	43

DISCUSSION

The present study reports consistent findings that accord with the model proposed in the extended-U. (Hancock & Warm, 1989) The current results indicate that cognitive performance change varies between an apparent insensitivity to stress during initial stress exposure to the catastrophic collapse that has been described during real-world disasters (Morgan, 1995; Leach, 2004). Assessment of cognitive functioning after one week of psychological and physical stress indicated slower than baseline responding with minimal accuracy change. Completion of items was self-paced, and soldiers elected to increase the

time they expended on each item in order to maintain accuracy as cognitive functioning declined. Had items been externally paced, increased error rates may well have occurred as the available time approached its limit.

Memory tasks deteriorated most. This pattern differs from previous studies which found Logical Reasoning was most sensitive to the stress of physically demanding simulated combat (Harris & Hancock, 2003). Replication of Class 1 changes in Class 2 indicates a strong consistency in the cognitive processing changes produced by SERE training. The difference in the pattern of cognitive tasks affected by stress suggests that the effects of physical and psychological stress do differ. Physical effort, discomfort, and sleep deprivation are high during simulated combat, but only rarely is the level of psychological threat elevated. In contrast, the SERE training had lower physical demands but included significant levels of psychological stress.

The appearance of decrements in complex cognitive task performance at the point at which SRT increased by approximately 20 ms is consistent with our previous studies that have found SRT diagnostic and that complex cognitive performance changes can be predicted by SRT changes (Harris et al., 2005; Harris & Hancock, 2003). The consistent appearance of cognitive performance deficits when SRT increases reach 20 ms suggests that SRT may provide an effective fitness for duty test. Memory deficits are consistent with cognitive changes during SERE training (Morgan, et al., 2004) and suggest that post SERE assessments are indicative of changes during stress. Furthermore, changes in the present study indicate that the information processing deficits found in the present study are likely to be present during stress.

The cognitive performance return to baseline in the present study further delineates the speed of recovery of cognitive functioning in military personnel but the minimum time for complete recovery still needs to be determined. Almost complete recovery occurred here after 12 hours of rest, but three hours was insufficient for complete recovery in an earlier group of Navy SERE students (Harris et al., 2005). We conclude that optimal recovery time lies between this three and twelve hour interval. Future research is needed to determine the relationship between cognitive function changes and operational performance.

ACKNOWLEDGEMENTS

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, P.A. Hancock, Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

REFERENCES

Benedetto J, Harris, W.C., & Goernert P.N. (1995). Assessing gender differences and norm data on a cognitive performance measure. *Proceedings of the Human Factors and Ergonomics Society, 39*, 968.

DeJohn, C.A., & Reams, G.G. (1992). *An analysis of a sustained flight operation training mission in Navy attack aircraft.* NAMRL-1370, Naval Aerospace Medical Research Laboratory, Pensacola, FL.

Hancock, P.A., & Szalma, J.L. (2003) Operator stress and display design. *Ergonomics in Design*, *11* (2), 13-18.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519-537.

Harris, W.C., Hancock, P.A. & Harris, S.C. (2005) Information Processing changes following extended stress. *Military Psychology, 17* (2),115-128.

Harris, W.C., & Hancock, P.A. (2003). Field assessment of cognitive performance under stress. *Proceedings of the Human Factors and Ergonomics Society, 47,* 1953-1957.

Leach J. (2004). Why people 'freeze' in an emergency: Temporal and cognitive constraints on survival responses. *Aviation Space and Environmental Medicine, 75*, 539-542.

Morgan, III C.A., Hazlett, G.A., Rasmusson, A., Zimolo, Z., Southwick, S.M., & Carney, D.S. (2004). Relationships among plasma dehydroepiandrosteron sulfate and cortisol levels, symptoms of dissociation and objective performance in humans exposed to acute stress. *Archives of General Psychiatry, 61*, 819-825.

Morgan W.P. (1995). Anxiety and panic in scuba divers. *Sports Medicine, 20* (6), 398-421.

Reeves D.L., Winter K.P., LaCour S.J., Raynsfor M., Vogel K., & Grissett J.D. (1991). *The UTC-PAB/AGARD stress battery: User's manual and system documentation.* Pensacola, FL: NAMRL; Special Report 91-3.

Scales, R.H. (2003) *Yellow smoke: The future of land warfare for America's military*. Rowman & Littlefield. Boulder, CO.

Wickens C.D., & Flach J. (1988) Human information processing. In: E. L. Wiener and D.C. Nagel (Eds.) *Human factors in aviation.* (pp 111-155) New York: Academic Press.

TESTING THE EFFICIENCY AND EFFECTIVENESS OF SOLDIER PERFORMANCE UNDER INCREASING TASK DEMAND

K.A. Burke, J. Szalma, T. Oron-Gilad, A. Duley, and P.A. Hancock
University of Central Florida
Orlando, FL

Combat is inherently a demanding situation which may increase stress, heighten arousal, and increase anxiety. The Processing Efficiency Theory (PET) was specifically developed to account for how anxiety influences performance. Therefore, the PET provides a potential theory to explain the positive and negative changes in performance in a combat environment. This study is the first attempt to examine PET in the military domain. Using the Small Arms Simulator Testbed (SAST), we investigated the relationship between processing efficiency and performance, considering mental workload, stress, and anxiety effects. Shooting performance effectiveness measures were target acquisition and friend/foe discrimination. Changes in processing efficiency were manipulated by varying degrees of working memory demand and sustained information transfer. The results indicated that shooting performance and processing efficiency, as well as mental workload demands, decreased as the global demand of both tasks increased. Further analyses for anxiety and stress and future directions are discussed.

INTRODUCTION

Combat is inherently a demanding situation which may increase stress, heighten arousal, and increase anxiety especially when the outcome is important and uncertain (Jones & Swain, 1995). Additionally, in a combat environment, maintaining a high level of performance, both of cognitive skills (i.e., decision making tasks) and motor skills (i.e., shooting accurately) is critical and may mean the difference between life and death. Although the relationship between anxiety and performance has been investigated extensively (Eysenck & Calvo, 1992; Yerkes & Dodson, 1908), the mechanisms underlying the facilitative or debilitative effects of anxiety on performance are not yet clearly understood. The Processing Efficiency Theory (PET; Eysenck & Calvo, 1992) was specifically developed to account for how anxiety influences performance. Traditionally, anxiety has been considered detrimental to performance, however a significant body of literature suggests that anxiety has both facilitative and debilitative functions (e.g., Jones, Hanton, & Swain, 1994; Jones & Swain, 1992; Jones, Swain, & Hardy, 1993). The PET (Eysenck & Calvo) seeks to explain how individual differences in trait anxiety interact with variations in state anxiety and attentional mechanisms to affect performance. The PET has primarily been applied to the examination of test anxiety (Eysenck, 1996), but the theory is broad enough to have wide application in other domains. In spite of its potentially broad applicability, the theory has not been empirically validated in other domains (Woodman & Hardy, 2001) with the exception of one study in the sport psychology domain (Murray & Janelle, 2003). As far as we know, our study is the first to investigate the PET in the military domain and in a high fidelity environment. The SAST II (Small Arms Simulator Testbed II) is a high fidelity simulation allowing for superior assessment of shooting performance in simulation. In this study the SAST II was utilized to test hypotheses based on PET. In many instances performance on a primary task may be maintained in the face of anxiety or stress only at the cost of increased effort. That is, performance may remain effective but less efficient. Primary task performance may therefore mask detrimental effects of anxiety and stress.

The aim of the current study was to identify changes in processing efficiency under varying degrees of working memory demand and sustained information transfer and to examine the interaction between task-related stress.

METHOD

Participants

Thirty-eight participants (mean age = 20.4, *SD* = 4.6, range = 18-43*)* were recruited on a voluntary basis from undergraduate psychology classes at the University

of Central Florida and were paid $22.50 for their participation (based on an hourly rate of $7.50).

Materials

Questionnaires. Participants completed several subjective questionnaires to measure individual differences including workload, stress, and anxiety. All subjective measures were administered via Inquisit, psychological data collection software, on a Dell Inspiron 5150.

Experimental Tasks

Primary Task. Target acquisition task and a Friend/foe discrimination task: two different types of stimuli/targets appeared on the screen (Figure 1): the non-target was an M-16 (friend), the target was an AK-47 (foe); both stimuli/targets were superimposed over the silhouette stimulus; the discrimination was based upon the weapon. The spatial location of each stimulus (targets/non-targets) was randomly generated to prevent the predictability of the appearance of a target.

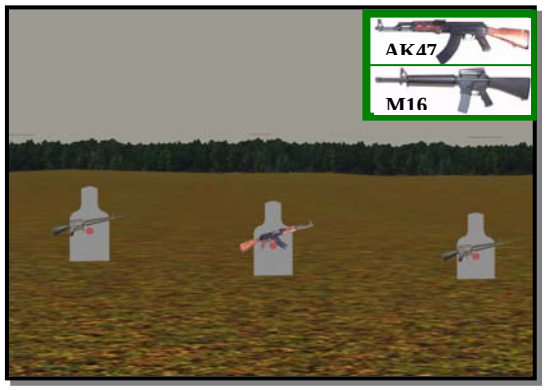


Figure 1. Screenshot from the SAST II displaying the two types of targets superimposed on the silhouette stimuli.

Secondary Task. Memory recall task: consisted of the auditory presentation of cue words and association words which simulated unit battalion numbers (e.g., 21, 23, 27) and locations (north, south, east, west), respectively, and were designed to provide varying demands on working memory. There were three secondary task conditions which varied according to working memory task demand, these being zero (control condition), and two or four items to be retained in memory. During execution of the primary task the participant was presented with the secondary task. The participants responded to the secondary task stimuli via three triggers mounted to the weapon. One trigger was pressed in response to the presentation of a valid cue word and true association word, the second trigger was pressed in response to the presentation of a valid cue word and false association word, and the third trigger was pressed in response to an invalid cue word.

Apparatus

Small Arms Simulator Testbed II (SAST II). (Naval Air Warfare Center Training Systems Division, 2000) is a single user research testbed for the evaluation and study of both current and new weapon concepts. The SAST controlled the presentation of targets (target type, location, time of onset, and exposure time), and recorded weapon aim-point data and shot results from an M16 rifle equipped with an infrared emitting diode and collimator lens. Weapon recoil was simulated with an electromechanical recoil system and weapon sounds were simulated via digital audio surround-sound (as shown in Figure 2).



Figure 2. The SAST II facility including the weapon and the visual display.

LabVIEW 7.0. (National Instruments, 2004). LabVIEW is a graphical development environment for creating flexible and scalable test, measurement, and control applications rapidly and at minimal cost. For the current study, LabVIEW was used to integrate the SAST II system, and the presentation of the secondary task.

Design Overview

The dependent measures of interest were: 1) response time for the primary task (mean response time participants required to pull the trigger on the weapon after the appearance of the targets); 2) response time for the secondary task (mean response time participants required to press a trigger in response to the secondary

task stimuli); 3) shooting accuracy; 4) accuracy on the secondary task; and 5) subjective measures.

The experiment included the following within-participant factors: 1) presentation of secondary task (present, not present); 2) increasing levels of working memory demand tasks (recognition/recall; three levels); and 3) friend/foe discrimination task (half of the trials included a target (foe), half of the trials did not include a target.

Procedure

Each participant was randomly assigned to one of the three working memory presentation orders, with the constraint that an equal number of participants were tested in each order. For each participant, the sand bags enabling a seated, supported firing position were adjusted for differences in height, the weapon was calibrated for accurate tracking and aiming vectors, and the triggers attached to the weapon (to allow the participant to respond to the secondary task stimuli) were adjusted for handedness and finger position. The participants received training and practice so that they would become familiar with firing the weapon, identifying the targets, and the secondary task.

Each session (approximately three hours) consisted of administration of the subjective measures, training and practice on both the primary and secondary tasks, and of the presentation of the three conditions of the secondary task. Each participant completed shooting scenarios for each condition of the secondary task (each working memory demand condition) which consisted of four blocks of 16 trials totaling 64 trials per working memory condition. Each trial consisted of the simultaneous presentation of one of two types of target stimuli (target, or non-target) and in half of the trials, the presentation of the secondary task. The trials were organized such that each participant was presented with an equal number of trials which included: the secondary task with an enemy target, secondary task without an enemy target, no secondary task with an enemy target, and no secondary task without an enemy target.

The presentation of trials within each working memory demand condition was randomized so that each condition was presented to each participant an equal number of times. Additionally, the order of presentation of each working memory demand condition (0, 2, and 4) was counterbalanced. Following each block, participants completed the RSME and between conditions, participant received a five-minute break.

RESULTS

Primary Task Performance. A two-way 2(secondary task present or not) X 3(working memory (0,2, and 4)) within-participants repeated measures analysis of variance (ANOVA) was conducted on performance in the friend/foe discrimination primary shooting task. The independent variables consisted of presentation of the secondary task in which the battalion number and location combinations were presented to the participant and varying levels of working memory demand. The dependent variable was the percent correct (number of times the participant correctly identified and hit the target) at each level of working memory demand. The main effect of presence of the secondary task and working memory demand on shooting performance was not statistically significant, Wilk's Λ = .98, F (1, 36) = .38, p = .69, partial η^2 = .11, and Wilk's Λ = .99, F(1, 36) = .03, p = .87, partial η^2 = .001, respectively. However, as depicted in Figure 3, in working memory conditions 0 and 2, participants were more accurate in the shooting task when there was no secondary task than when the secondary task was present. This was not the case in working memory condition 4; possibly due to the combination of practice and fatigue effects.

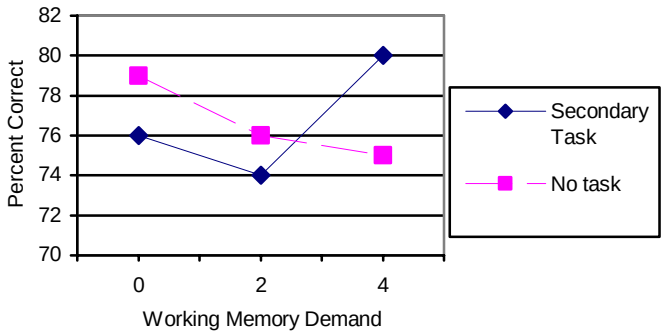


Figure 3. Relationship between presence of the secondary task and working memory demand.

Secondary Task Performance. A two-way 2(primary task target present or not) X 2(working memory demand (2, 4)) within-participants repeated measures ANOVA was conducted on secondary task performance (percent correct-- number of times the participant correctly recognized and responded to the presentation of the battalion number and azimuth location). There was a significant main effect of presence of a target and working memory demand, Wilk's Λ = .61, F (1, 18) = 11.47, p = .003, partial η^2 = .39, and Wilk's Λ = .76, F(1, 18) = 5.61, p = .03, partial η^2 = .61, respectively. These results show that participants were significantly more accurate on the recognition task when there was not a

target present than when a target was present and when the working memory demand level was two (see Figure 4).

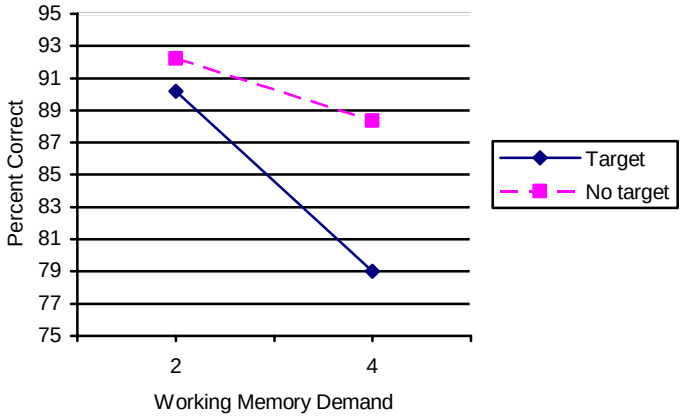


Figure 4. Relationship between the presence of a target and working memory demand.

Perceived Workload. A 3(working memory demand) X 4(block) within-participants repeated measures ANOVA was conducted on the subjective ratings of mental workload (RSME). There was a significant main effect of working memory demand and block, Wilk's Λ = .35, F (2, 31) = 29.32, p = .001, partial η^2 = .65, and Wilk's Λ = .55, F(3, 31) = 8.08, p = .001, partial η^2 = .45, respectively. Specifically, participants reported that a higher level of mental effort was required as a function of working memory demand; as the working memory demand increased, the ratings of mental effort increased (see Figure 5). Additionally, and consistent with our hypotheses, participants reported that a lower level of mental effort was required as a function of time in each working memory demand condition.

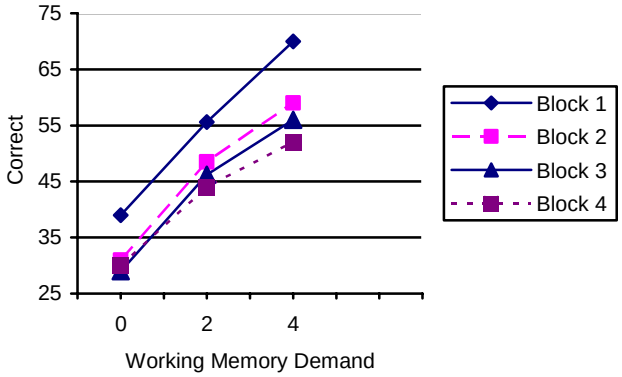


Figure 5. Relationship between subjective mental effort and working memory demand.

DISCUSSION

This study is the first to investigate processing efficiency in a military context and in a high fidelity simulator. The use of a high fidelity simulator is important in this case because the target aiming and target acquisition task are performed using a real weapon which adds to the physical and the hand-eye coordination demands of the task. As can be seen from the results there are changes in performance for both the primary and the secondary task as the global demands change, clearly indicating that it is not the case of one task dominating the other. However, contrary to our hypothesis, when the task was most demanding (both primary and secondary tasks were present in the highest working memory demand condition), participants' were more accurate on the shooting task when the secondary was present than when there was no secondary task. One possible explanation for these results is that there were combined practice and fatigue effects. Hence, the variability among participants increased, and possibly the analysis of the anxiety data could contribute to the understanding of these changes. Additionally, because participants were significantly less accurate on the recognition task in the same high demand condition, it is possible that participants stopped attending to the secondary task and focused all of their attention on the primary, more salient shooting task. This would be consistent with the Hancock and Warm (1989) model indicating that participants were no longer in their comfort zone and were unable to maintain an optimal level of performance on both tasks. Furthermore, such pattern could also be explained using Hockey's (1997) compensatory control model. Specifically, it may be that as attentional demand increased participants regulated their effort downward on the secondary task to free resources required to successfully engage the primary task.

Furthermore, the mental workload measures correlate with this global demand, indicating a performance-workload association (Hancock, 1996; Yeh & Wickens, 1988). Further investigations of this relationship are necessary. Current efforts are underway to examine the hypothesized effects of trait anxiety on performance and perceived workload. In addition, the other form of workload assessment, physiological measurement, will be examined to test whether there is physiological evidence that participants reduced their effort in the face of high working memory demand. These analyses will provide a more coherent understanding of the results. Future studies will be able to refine these findings and better explain the tradeoff in performance as task demand increases.

ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621, Peter A. Hancock, Principal Investigator.

REFERENCES

Baddeley, A.D. (1986). *Working memory.* Oxford, UK: Clarendon Press.

Eysenck, M.W. (1996). Anxiety, processing efficiency theory and performance. In W. Battmann & S. Duke (Eds.), *Processes of the molar regulation of behavior* (pp. 99-104). Scottsdale, AZ: Pabst Science Publishing.

Eysenck, M.W., & Calvo, M.G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*, 409-434.

Hockey, R.G. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1-3), 73-93.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*(5), 519-537.

Jones, G., & Swain, A.B.J. (1992). Intensity and direction dimensions of competitive state anxiety and relationships with competitiveness. *Perceptual and Motor Skills, 74*, 467-472.

Jones, G., Swain, A.B.J., & Hardy, L. (1993). Intensity and direction dimensions of competitive state anxiety and self-confidence. *Journal of Sports Sciences, 11*, 533-542.

Woodman, T. & Hardy, L. (2001). A case study of organizational stress in elite sport. *Journal of Applied Sport Psychology, 13*(2), 207-238.

Yeh, Y., & Wickens, C.D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30*(1), 111-120.

Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology, 18,* 459-482.

Zijlstra, F.R.H. (1993). *Efficiency in work behavior. A design approach for modern tools*. PhD thesis, Delft University of Technology. Delft, The Netherlands: Delft University Press.

Zijlstra, F.R.H. & Van Doorn, L. (1985). *The construction of a scale to measure perceived effort.* Delft, The Netherlands: Department of Philosophy and Social Sciences, Delft University of Technology.

EVALUATION OF THREAT BY POLICE OFFICERS: INITIAL FINDINGS

Karen A. Delos Santos, Shawn C. Stafford, James L. Szalma,
Tal Oron-Gilad & P.A. Hancock
University of Central Florida
Orlando, Florida

Police officers' threat perception of static images was examined using images reflecting the range of five threat categories on which police officers are trained. Thirteen experienced officers from a police departments in the southeastern United States participated in this study. Officers rated their perceived threat level for 110 images that were presented to them on a laptop computer. Each of these images was rated twice by each officer. Officers used all five categories to rate the stimuli, and their responses to the extremes (images rated as 1 or 5) were faster than responses to more ambiguous stimuli in the other categories. These results were generally consistent with predictions based on Fuzzy Signal Detection Theory. Further studies will evaluate performance with these images in the context of a signal detection task. Once fully developed, this tool could be used to evaluate new recruits' decision-making process before given the green light to carry a badge. These assessments could also be used as a modified training tool for experienced officers if the stimuli were to be placed in a semi-immersive environment.

INTRODUCTION

A major task Police Officers confront on a daily basis is situation assessment and subsequent determination of the level of threat posed to either themselves and/or to the public they are sworn to protect. There are situations in which they may perceive no threat or other situations in which they may perceive total threat. These perceptions will elicit different responses based on the officer's training and the protocol they were instructed to follow given different threat levels. The way an officer perceives a situation's threat level will determine the manner in which he or she will react, as well as how quickly they will make a decision. For instance, an individual running toward an officer, or an individual reaching into his pocket, represents a rather ambiguous situation the officer must assess accurately in a short period of time. The response to a determined threat level can range from little to no reaction (visual acknowledgement or complete dismissal) to total threat (use of deadly force to subdue the threat). This range of response makes it imperative that the officer correctly identifies the situation's appropriate threat level. To be considered a 'successful' response, the threat level must not greatly contradict the officer's response. This decision-making task can be conducted many times over the course of a work-shift depending on the officers' assignment and work-environment. The consequences of their decision could mean the difference between life and death for them, the citizens of the community or the suspect in question.

In spite of the national and international importance of threat detection in law enforcement, little systematic research into the perceptual

processes underlying these decisions has been done. This study represents an initial effort to develop threat stimuli for such a systematic research. Thus, we provided officers with images sampled from the range of threatening situations they may face daily in order to examine how they categorized these events and whether their response times to those events varied according to threat rating. Finally, we also sought to determine the degree to which officers were consistent in their ratings of images. Each officer therefore experienced each image two times.

Police officers typically evaluate threat level of events using five categories, with the lowest (1) being no threat and the highest (5) being 'total threat' (i.e., life threatening). Threat level 1 corresponds to officer presence and often requires no interaction with the subject or subjects. Threat level 2 would require verbal commands and possibly non-aggressive physical touch leading subjects to another location. Threat level 3 would require controlling force with moderate physical exertion. Threat level 4 would often require incapacitating force endangering the life of the suspect and police officer. Threat level 5 often could require deadly force at the discretion of the police officer. That is, police officers in threat level 5 are not required to use deadly force, although such a response is justified. We adopted this threat level scheme for this study, since it accords with both the training police receive and the way they actually assess events as they occur in the course of a shift.

This study aims to develop and validate stimuli that could potentially investigate what police officers perceive as threat and the specific threat level they assign to each situation. With validated stimuli, many different avenues for research can be explored, such as police stress, vigilance in task allocation, and workload sensitivity. Detailed assessments could be further used to evaluate new recruits decision-making process before they are given the green light to carry a badge. These assessments could also be used as a modified training tool for experienced officers if the stimuli were to be placed in a semi-immersive environment. Finally, the results of the current study can be used to derive mapping functions for use in Fuzzy Signal Detection Analyses (Parasuraman, Masalonis, & Hancock, 2000), which will permit a more accurate assessment of threat detection performance in complex real-world situations that routinely face law enforcement professionals.

METHOD

Thirteen police officers were chosen at random from a police department from the southeastern United States. Police officers volunteered their time and were not offered any reward or incentive. Each participant signed an informed consent and was given a debriefing form at the conclusion of the session. There were 11 male and 2 female participants. Their ranks were: officer (10), sergeant (1), and lieutenant (2). They were all assigned to road patrol with several officers also assigned to specialized units (S.W.A.T. and K9 division). The average age of the officers was 32 (SD=7). The average years of service were 6.8 (SD=5.7). The officers were either assigned to mid shift (3 P.M.-11 P.M.) or night shift (11 P.M.-7 A.M.).

Each participant was asked to read an informed consent and sign the second page if they agreed to the provisions provided within. Participants then completed a short demographic questionnaire. The experiment was presented to each officer on a laptop computer. The participants read a series of directions presented on the computer prior to judging the stimuli. There were 110 different pictures presented which depicted scenes that are encountered by police officers. The pictures were taken from a several databases of police interactions, all of real world interactions depicting actual events (as shown in Figures 1 and 2). Each image was presented twice during the session. The order of presentation was randomized, with the restriction that the two presentations of an image

never occurred on consecutive trials. Each stimulus presentation was 3000 ms in duration. The police officers were instructed to respond using keys 1, 2, 3, 4, and 5 (1- no threat, 3-between no threat and total threat, and 5- total threat). Once the police officer responded, the next picture was presented.



Figure 1. Example of category 2.



Figure 2. Example of category 5.

RESULTS

To test for reliability between the officers' two ratings for each picture, a series of bivariate correlations were computed for all 110 stimuli. The average correlation between image presentations was .70, indicating a reasonable degree of consistency in categorization of images by the officers.

For each of the 110 stimuli, the mean response was computed. There were a total of 21 level 1 images across the participants, 32 level 2 images, 30 level 3 images, 16 level 4 images, and 11 level 5 images. Within each category of responses, the median response times were then calculated for each officer, and the mean of the median response times was computed across all 13 participants. The mean response time scores are shown in Table 1.

Table 1 – Category of threat, number of images associated with each category and the mean median response.

Threat Category	1	2	3	4	5	Total
No. of Images	21	32	30	16	11	110
Response Time (ms)	606.43	711.88	729.15	825.72	563.18	698.148
SD	131.08	219.80	207.81	444.01	95.31	250.026

An ANOVA (threat level X reaction time) indicated a significant main effect for threat category, $F(4, 105) = 2.73$, $\eta^2 = .762$, $p = .026$. Post hoc pair-wise comparisons revealed that response times to stimuli categorized at level 4 (M = 825.72, SD= 444.01) were significantly *slower* than those to images categorized as category 1 (M = 606.43, SD= 131.08). Response times to images categorized at threat level category 4 were also significantly slower than those to stimuli categorized at level 5 (M=563.18, SD = 95.31).

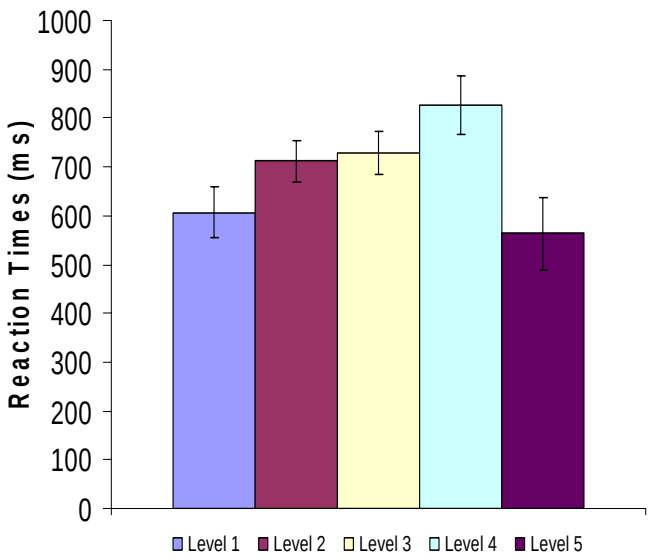


Figure 3. Reaction time by threat level category (error bars are standard errors)

DISCUSSION

In general, the results of this study accorded with expectations. Officers used all five categories to rate the stimuli, and their responses to the extremes were faster than responses to more ambiguous stimuli in the middle categories (2-4). The latter results accord with those reported by a prior experiment in which observers were required to rate the degree to which a stimulus was a signal on a scale from 1 to 7 (Stafford, Szalma, Hancock, & Mouloua, 2003). In that study, Stafford et al. (2003) observed that responses were slower to stimuli in the middle categories, where the level of the stimulus is uncertain for the participant. The response time effects observed in both experiments are in agreement with the tenets of Fuzzy Signal Detection Theory (FSDT; Hancock, Masalonis, & Parasuraman, 2000; Parasuraman, Masalonis, & Hancock, 2000). FSDT represents a fusion of fuzzy set theory and signal detection theory (SDT), in which the mutually exclusive categories used in SDT (i.e., signal, non-signal) are replaced by categories in which joint membership is possible. This model allows for the quantification of real-world stimuli that possess properties of signal and non-signal to various degrees. An implication of FSDT is that the middle of the stimulus dimension should be the point of greatest uncertainty, since at that point the stimulus possesses the properties of both signal and non-signal to equal degrees. The preliminary results presented in our study are generally consistent with this notion. Subsequent research will validate use of these stimulus categories with a larger sample of officers for the purpose of employing these stimuli for experimentation using FSDT methodology. It is believed that FSDT will prove as a valuable tool in determining the threat level discriminations and responses as they relate to police officers decision making and execution of the correct responses.

ACKNOWLEDGMENT

The view expressed in thesis work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-0621, Peter A. Hancock, Principal Investigator

REFERENCES

Hancock, P.A., Masalonis, A.J., & Parasuraman. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science, 1*, 207-230.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42, 636-659.*

Stafford, S.C., Szalma, J.L., Hancock, P.A., Mouloua, M. (2003). Application of Fuzzy Signal Detection to Vigilance: The effect of Criterion Shifts. *Proceedings of the Human Factors and Ergonomics Society 47,* 1678-1682, Santa Monica, CA: Human Factors and Ergonomics Society.

Training for Vigilance:

The Effect of KR Format and Dispositional Optimism and Pessimism on Performance and Stress

James L. Szalma and Peter A. Hancock

Department of Psychology and Institute for Simulation and Training

University of Central Florida

William N. Dember and Joel S. Warm

Department of Psychology

University of Cincinnati

Address correspondence to: J.L. Szalma, MIT^2 Laboratory, 3100 Technology Parkway, Ste 333, University of Central Florida, Orlando, FL, 32826; jszalma@mail.ucf.edu

Running Head: Optimism, Pessimism, and Vigilance

Keywords: Optimism; Pessimism; Individual Differences; Vigilance; Sustained Attention; Stress

Training for Vigilance:

The Effect of KR Format and Dispositional Optimism and Pessimism on Performance and Stress

ABSTRACT

This study investigated the impact of knowledge of results (KR) format on the performance and stress associated with a vigilance task. Also examined was the effect of the interaction of KR-format (Hit-KR, False Alarm-KR, Miss-KR, and a composite of all three formats) with dispositional optimism and pessimism on performance outcome and reported stress state. Hypotheses based upon a theory of feedback intervention were tested. KR regarding correct detections and the composite-KR (KR regarding correct detections, false alarms, and missed signals) enhanced perceptual sensitivity. However, False Alarm-KR and Miss-KR did not. Contrary to expectations based on the theory, performance was unrelated to the traits across all KR conditions. However, the effects of KR-format on self-reports of stress depended on the individual's level of pessimism and optimism. In addition, KR format and personality affected the multiple dimensions of stress state in different ways.

INTRODUCTION

Tasks demanding vigilance or sustained attention require observers to monitor displays for the appearance of infrequent signals for action over prolonged periods of time. The central issue, in almost sixty years of vigilance research, has been the decline in performance with time on watch termed the vigilance decrement (See, Howe, Warm, & Dember, 1995). Although vigilance has been and remains an important part of many contemporary work environments, it has assumed significantly greater importance with the introduction of many 9/11 mandated security systems such as by the Transportation Security Administration's (TSA) need to screen 100% of all baggage at all U.S. airports (Harris, 2002; see also Hancock & Hart, 2002). The advent of these and many similar security requirements has increased the need for effective training procedures for system operators. In vigilance, training has traditionally employed feedback in the form of knowledge of results (KR) regarding an operator's accuracy expressed in terms of correct detections, false alarms, or missed signals. KR has been shown to improve performance on sustained attention tasks (for a review, see Warm & Jerison, 1984). However, there is experimental evidence that the form of KR can impact its effectiveness. For example, Dittmar, Warm, and Dember (1985) asked participants to monitor increments in a vertical line presented on a video display terminal (VDT) and provided observers with feedback, in the form of a 1000 Hz tone, regarding their correct detections, false alarms, or missed signals. They observed that Hit- and False Alarm-KR enhanced perceptual sensitivity (A') and attenuated the vigilance decrement but Miss-KR had no such effects. They explained these findings in terms of the information provided by each KR format. Hit- and False Alarm-KR provide information regarding responses to specific stimuli but miss KR is not linked to the observer's noticing of any particular characteristic of a signal event or to the observer's subsequent response.

Although Dittmar and her colleagues (1985) reported performance effects of KR format, they did not examine the stress effects associated with the use of KR. Several experiments have reported that vigilance tasks are stressful. Observers report feeling more boredom, strain, irritation, distress, and less task engaged after the task than prior to its start (e.g., see Szalma et al., 2004; Warm, 1993). In addition, research has indicated that manipulations of task characteristics that influence performance also influence individual subjective state (e.g., Szalma et al., 2004; Temple et al., 2000; and see also Warm, Dember, & Hancock, 1996). It is feasible therefore that the format in which KR is provided to observers can influence their stress state in addition to influencing their performance. Given that Dittmar et al. (1985) observed performance differences as a function of KR format, and the associations between performance and subjective state (Warm, Dember, & Hancock, 1996), one might expect that Hit- and False Alarm-KR would be associated with reduced stress symptoms and that Miss-KR would be associated with greater stress, relative to a no-KR control condition. That is, the effect of KR format on stress should mirror the effects previously reported by Dittmar and her colleagues in regard to performance. Previous studies using a composite format have shown that used together the three forms of KR enhance performance (e.g., Szalma, Miller, Hitchcock, Warm, & Dember, 1998; Szalma, Hancock, Warm, Dember, & Parsons, 2005) but these studies did not examine the effects of composite KR on observers' stress states. It may be that providing all three forms of KR will show the greatest performance enhancement and reduction in stress, since all information regarding detection accuracy is explicitly provided. One goal for the current experiment was to test these propositions.

Individual Differences in Performance and Stress Response: Feedback Intervention Theory (FIT)

From existing evidence, it appears highly likely that the effects of KR format on performance and stress associated with vigilance will depend upon the characteristics of the individual. Indeed, it has been long recognized that general individual differences among observers can influence performance in vigilance (see Berch & Kanter, 1984; Davies & Parasuraman, 1982). Recent evidence has confirmed that responses to the stress of sustained attention also depends on characteristics of the individual observers (e.g., Helton, Dember, Warm, & Matthews, 1999; Szalma, 2002a). Such findings are consistent with transactional models of stress that emphasize the importance of appraisal processes in stress response (Lazarus & Folkman, 1984; Matthews, 2001).

An individual differences variable that has recently emerged as an influential factor on stress responses to sustained attention is dispositional pessimism/optimism. Helton, Dember, Warm, and Matthews (1999) investigated the effect of false failure feedback on the performance and stress of both pessimists and optimists in a vigilance task. Although false failure feedback did not significantly impact performance, individuals high in pessimism and low in optimism exhibited steeper vigilance decrements across conditions relative to individuals low in pessimism and high in optimism. Helton and his colleagues also observed that pessimistic individuals reported changes in stress related parameters such as greater tension, lower hedonic tone (i.e., a more negative mood state), and less confidence relative to optimists. These findings suggest that the high stress associated with vigilance may be moderated by an observer's relative level of pessimism/optimism. Further evidence for the influence of pessimism and optimism on observers' subjective stress states was reported by Szalma (2002a) who found that higher

degrees of pessimism were associated with greater *Distress* and lower *Task Engagement* after a vigilance task. Szalma (2002a) also reported that pessimism predicted greater reliance on emotion-focused coping, consistent with prior research indicating use of this coping strategy by pessimistic individuals. Optimism, in contrast, was less predictive of stress and coping.

In regard to individual differences in the effectiveness of KR for vigilance, predictions can be made based upon a Theory of Feedback Intervention (FIT), derived from a meta-analytic study of the feedback intervention literature (Kluger & DeNisi, 1996). FIT serves as a general framework for understanding the processes that influence feedback effectiveness. Drawing on control theory (see Carver & Scheier, 1981; 1998), FIT assumes that behavior is regulated by comparison of feedback to goals or standards that are hierarchically organized. Two further assumptions of this theory are that 1) only those feedback interventions that receive attention can influence performance and 2) that feedback interventions have the capacity to alter the locus of attention within the goal hierarchy. Although hierarchies can vary in complexity, Kluger and DeNisi (1996) describe three general levels relevant for feedback effects. At the lowest level are *task-learning processes* consisting of 'hypotheses' for improving performance that individuals test by utilizing feedback. Attention to this level of the hierarchy constitutes attention to the details of focal task performance. These processes are regulated by the next highest level in the hierarchy, *task-motivation processes*, which serve as standard or comparators for the task-learning processes. At the highest level are *'meta-task' processes* involving the self and general but highly salient goals. Specifically, *meta-task processes* relate task outcomes to higher level goals that are important to the individual. Kluger and DeNisi (1996) argued that most affective reactions to feedback interventions, when they occur, result from evaluation of feedback with respect to salient self goals specified at the meta-task level.

According to FIT, feedback is effective in improving performance only to the extent that it serves to focus attention on the task (i.e., *task-motivation* and *task-learning processes*). Thus, an FIT interpretation of the results of Dittmar and her colleagues (1985) would be that the Hit- and False Alarm-KR directed attention to task-learning processes (e.g., learning the nature of a signal/non-signal) and perhaps task-motivational processes (e.g., see Warm, Kanfer, Kuwada, & Clark, 1972; Warm, Epps, & Ferguson, 1974), while Miss-KR failed to adequately direct attention to relevant task characteristics. Miss-KR may fail to direct attention appropriately because it is not linked to a specific, overt response from the observer, so that observers are unable to use the feedback to associate their response to the stimulus characteristics (and see Szalma et al., 2005). Moreover, one of the propositions of FIT (see Proposition 5, Kluger & DeNisi, 1996, p. 269) is that feedback cues that match salient personal goals associated with a given personality trait will direct attention to meta-task processes and thereby debilitate performance. This may subsequently induce stress effects in such individuals. Kluger and DeNisi (1996) suggested a general mechanism for these effects whereby individuals' traits influence their ability to redirect their attention to the details of the task in response to feedback. They proposed that this occurs via an inhibition mechanism that protects working memory space from competing demands imposed by the goal hierarchy, and that traits such as the tendency to blame oneself for failures may exert debilitative effects on feedback by interfering with this inhibition mechanism. Hence, individual differences relevant to an observer's response to feedback may influence the effectiveness of training with feedback interventions such as KR.

Given the propositions of Kluger and DeNisi (1996), traits such as pessimism and optimism, which impact general expectancies regarding performance and affective responses to

performance success and failure (see Chang, 2002), should influence the effectiveness of KR in improving vigilance. Specifically, the negative outcome expectancies associated with pessimism may serve to interfere with these inhibitory processes and thereby reduce KR effectiveness, particularly for 'negative' KR (i.e., False Alarm-KR and Miss-KR), since individuals high in pessimism tend to appraise events more negatively than those low in pessimism. Such feedback may, in Kluger and DeNisi's (1996) terms, 'release' the pressure on the regulatory system to eliminate the feedback-standard discrepancy at the task level. Attention is then redirected toward the self and managing the negative affect associated with pessimistic appraisals of the task. In contrast, the more positive expectations of individuals high in optimism would result in feedback directing their attention toward more task-based processes, thereby improving performance and reducing stress symptoms. These effects may be weak, however, since optimism has been shown to be a weaker absolute predictor of other measures than pessimism (Szalma, 2002a; see also Hummer, Dember, Melton, & Schefft, 1992). The results of Helton et al. (1999) in regard to false failure feedback sit contrary to these hypotheses, as failure feedback would be expected to impair performance of pessimists by inducing negative appraisals and directing their attention toward dealing with their subsequent negative affect. However, the absence of significant KR effects in their study may be due to the fact that feedback was not veridical and was provided at the end of a block of trials (i.e., summary KR). Veridical feedback provided after each response (or failure to respond when appropriate) provides many more opportunities for the personality traits to differentially exert their influence on performance and stress state across feedback conditions.

The present study was thus also designed to test these propositions by examining the impact of individual differences in optimism and pessimism on performance and stress in

vigilance, and the effectiveness of KR in facilitating performance and reducing stress. Specifically, individuals higher in pessimism should exhibit greater performance impairment and report higher levels of stress symptoms after the vigil. Based on FIT it is hypothesized that veridical feedback emphasizing failure immediately following an error (i.e., Miss- or False Alarm-KR) should debilitate performance and increase stress for observers higher in pessimism (due to a shift in focus from the task toward the self) relative to a no-KR control condition. In the composite condition the effects of negative KR may be offset by the provision of Hit-KR. Thus, participant disposition should interact with feedback type, such that performance and stress effects associated with high pessimism should manifest themselves primarily in the context of False Alarm- and Miss-KR. In contrast, individuals high in optimism may respond positively to Hit-KR, although the findings of Szalma (2002a) and previous research on optimism and pessimism (Hummer et al., 1992) indicate that these effects may be weaker than those associated with pessimism.

EXPERIMENTAL METHOD

Experimental Participants

Fifty-five male and fifty-five female undergraduates from the University of Cincinnati participated in this experiment in exchange for course credit. Participants ranged in age from 18 to 30 years old, with a mean of 20.0 years. All participants had normal or corrected-to-normal vision and were free of any known hearing impairments.

Feedback Conditions and Task Characteristics

Twenty-two participants were assigned at random to each of the following five groups, with the only restriction being that the groups were equated for sex: Hit-KR; observers received feedback regarding correct detections. False Alarm-KR; observers received feedback when they made errors of commission. Miss-KR; observers received feedback when a signal appeared on the screen and they failed to respond to it; Composite-KR; observers received all three forms of feedback; No-KR; observers received no feedback regarding performance quality and acted as the comparison control group. The task required monitoring a simulated chemical process display presented on a video display terminal (VDT) for patterns of deviation that would lead to an uncontrolled reaction (see Figure 1). The display consisted of Cartesian co-ordinates (formed by 3.2 cm green line segments) representing the state of four system variables, each of which was represented as a point on one of the axes. These points were connected via green line segments, which formed a polygon that changed shape according to changes in system variables. Representative stimuli are displayed in Figure 1. The variables to be monitored were the pressure, temperature, and the flow rates in and out of the 'vat' of chemicals. Neutral events, which required no overt response from the observer, were cases in which these variables were in balance (see display (a) in Figure 1) or when their levels varied along one axis (e.g. pressure/temperature; see displays (b) and (c) in Figure 1) but not the other (e.g. flow rate; see displays (d) and (e) in Figure 1). There were five such events. Under one of these conditions the four vertices along the axes (each representing the level of a system variable) were .8 cm from the origin, resulting in a square shaped polygon. For the other four neutral events, one of the points moved to 1.2 cm from the origin, resulting in four different orientations of a 'kite' shape. Critical signals for detection were cases in which the flow rate-in did not equal the flow rate-out

and *both* temperature and pressure increased. Specifically, critical signals consisted of cases in which 1) the temperature, pressure, and flow rate-in points were moved from .8 cm from the origin to 1.2 cm from the origin; or 2) the temperature, pressure, and flow rate-out points were moved from the origin to 1.2 cm from the origin. Since critical and non-critical events were defined in terms of combinations of variable levels, the state of the overall system was conveyed via the shape of the polygon. Thus, neutral events were cases in which the figure displayed was either "square" or "kite" shaped, while critical signals were cases in which the quadrangle was not diamond or kite shaped.

Insert Figure 1 about here

Displays of this type are referred to as 'object' or 'volumetric' displays, and have been successfully employed by human factors researchers in the design of displays in the workplace (see Bennett & Flach, 1992, for a review; and see also Hancock & Szalma, 2003). An advantage of these displays is that they allow observers to assess the state of a multidimensional system by inspection of a geometric shape rather than by inspection of several separate gauges and recent evidence indicates that such displays can attenuate performance decrement, but that they do not reduce the stress of sustained attention (Szalma, 2002b). Note that the display employed in this study was not intended to simulate a specific real-world process, but rather to simulate the kind of displays that have been used to simplify the inspection of complex, multidimensional systems (Bennett, Nagy, & Flach, 1997).

Questionnaires

Observers' stress states were assessed with the Dundee Stress State Questionnaire (DSSQ, version 2.0; Matthews et al., 1999; see also Matthews et al., 2002), which is a multidimensional, self-report instrument for assessing transient mood states. Specifically, this version of the DSSQ is comprised of ten factor-analytically determined scales. These scales have been subjected to a secondary factor analysis, yielding three dimensions associated with stress state, *Distress*, *Task Engagement*, and *Worry*. Thus, increases in stress are associated with increased scores on the *Distress* and *Worry* scales and/or a decrease on the *Task Engagement* scale. Example items for *Distress* include ratings of the degree to which adjectives such as "tense" "relaxed" "happy" or "contented" describe the individual's current subjective state. Examples of items associated with *Task Engagement* include subjective state ratings of adjectives such as "energetic" "alert" or "vigorous," as well as ratings of statements such as "I expect the content of the task will be interesting" and "I am committed to attaining my performance goals." Example items associated with the Worry scale include ratings of statements such as "I'm reflecting about myself" and "I though about the difficulty of the problems."

Optimism and pessimism were assessed using the optimisim/pessimisim instrument (OPI) developed by Dember, Martin, Hummer, Howe, and Melton (1989). This instrument consists of a 56-item (18 items indicating optimism, 18 indicating pessimism, and 20 filler items) regarding the individual's attitudes and expectations. Scores on the OPI range from 18 to 72. Several studies using the OPI have indicated that optimism and pessimism are not simply polar opposites, but are partially independent (for a review see Dember, 2002). The OPI was administered pre-vigil, and the DSSQ was administered both pre- and post-vigil.

Procedure

Observers participated in a 30-min vigil divided into 6 continuous 5-min periods on watch. The display was presented for 500 ms every 1.5 sec (30 events/min), and critical signals for detection occurred twice/min (signal probability=.066) within each period of watch. KR was provided by a computer using a pre-recorded female voice to announce correct detections ("correct"), missed signals ("miss"), and errors of commission ("false alarm"). In the No-KR condition, the voice announced "saved" after each response in order to control for added auditory stimulation in the KR-conditions. Previous experiments using the female voice to deliver feedback found that this form of KR effectively enhanced performance (Szalma et al., 1998; Szalma et al., 2004).

Upon entering the laboratory participants surrendered their wristwatches and cell phones, read a brief description of the experiment, and signed a consent form. Participants completed the OPI, followed by the pre-task version of the DSSQ. The experimenter then read the instructions for the sustained attention task which were simultaneously presented on the VDT. Observers had no prior knowledge of the length of the vigil, other than that it would not exceed 90 minutes. At the end of the vigil participants completed the post-task version of the DSSQ and were then debriefed.

EXPERIMENTAL RESULTS

Performance Measures

Vigilance performance data are usually analyzed using signal detection theory (SDT) indices or their base component measures, hit and false alarm rates. However, in the present

experiment, as in many other vigilance experiments, a substantial proportion (37%) of the false alarm scores were zero across participants, periods, and KR group (e.g., Szalma et al., 2004; 2005; and see Davies & Parasuraman, 1982). Computation of parametric SDT parameters from zero false alarm scores requires adjustments to the data that can introduce error into the estimates (Craig, 1977; Davies & Parasuraman, 1982; Macmillan & Creelman, 2005). The difficulty in interpreting parametric SDT indices under these conditions led Craig (1977) to argue against the uncritical of these measures to vigilance data. He suggested (Craig, 1979) that nonparametric sensitivity indices (specifically, A') be used when parametric measures are inappropriate. Subsequent research established that among the available, non-parametric indices of response bias, β_d'' (Donaldson, 1992) was the most effective for vigilance, as this measure remained sensitive to non-perceptual manipulations and it is effective in estimating response bias at chance performance levels (See, Warm, Howe, & Dember, 1997). One advantage of non-parametric indices is that they are defined for cases in which there are many instances of zero false alarm rates, as in the present experiment. Correct detections and false alarms for each observer were therefore used to compute signal detection theory measures A' and β_d'' for each period of watch. A 5 (group) by 6 (period) analysis of variance (ANOVA) was computed for each of these respective performance measure. Mean sensitivity scores for observers in each KR condition are plotted as a function of period of watch in Figure 2.

Insert Figure 2 about here

Performance remained relatively stable across the watch period, and observers who received any form of KR achieved higher A' scores than those who did not receive KR. The ANOVA confirmed a significant effect for KR, $F(4, 105)=3.85$, $p<.01$, $\omega^2 = .02$, and a

significant effect for period on watch, $F(3, 318)=7.49$, $p<.001$, $\omega^2 =.05$. The latter effect is likely a reflection of the fluctuation in performance over the course of the vigil (M's for periods 1 through 6 were .87, .88, .85, .90, .86, and .88, respectively). The interaction between these factors lacked significance ($p>.05$). For all analyses involving repeated measures, Box's epsilon was used to adjust the degrees of freedom for violations of sphericity (see Maxwell & Delaney, 2004). Tukey HSD tests for the KR conditions revealed that relative to those in the No-KR control group (M=.81), observers who received Composite-KR (M=.92; Cohen's d=.84) or Hit-KR (M=.89; Cohen's d=.60) achieved significantly higher sensitivity scores. The four KR conditions did not differ significantly from one another in perceptual sensitivity ($p>.05$ in each case).

Mean response bias scores for observers in the five KR conditions are plotted as a function of periods of watch in Figure 3. The ANOVA indicated a significant effect for KR, $F(4, 105)=2.54$, $p<.05$, $\omega^2 =.009$, and for period on watch, $F(3, 291)=9.58$, $p<.001$, $\omega^2 =.06$. The interaction between these factors was not statistically significant ($p>.05$). Tukey HSD tests for the KR conditions revealed that observers receiving composite KR (M=.91) were significantly more conservative than those who did not receive KR (M=.49; Cohen's d=.80). All other comparisons failed to reach statistical significance ($p>.05$ in each case). The significant main effect for periods indicates that conservatism increased with time on watch (M's for periods 1 through 6 were .54, .67, .71, .73, .77, and .77, respectively). These results are consistent with prior research in vigilance, and are likely due to growing awareness by participants during the vigil that the signals to be detected were relatively rare (Craig, 1978).

Insert Figure 3 about here

Perceived Stress

Preliminary ANOVAs confirmed that participants across the KR conditions did not differ significantly in their pre-vigil *Distress*, *Task Engagement*, or *Worry* scores (p>.45 in each case). Pre- and post-DSSQ responses were therefore used to compute pre-post z-scores for the three DSSQ factors. One sample t-tests using the Bonferroni correction (α=.05) indicated that across all conditions observers reported significant increase in *Distress* (M=.66), $t(109) = 6.42$, p<.001, and significant declines in *Task Engagement* (M= -.86), $t(109)= -10.05$, p<.001, and Worry (M= -.46), $t(109)= -5.33$, p<.001. Thus, for two of the three dimensions of stress, the data indicate that the vigil itself was stressful. To examine KR effects, an ANOVA was computed for each stress scale. These analyses indicated that there were no significant differences among KR conditions for the *Distress* and *Worry* dimensions (p>.05 in each case), but that there was a significant KR effect for *Task Engagement*, $F(4, 105)=3.59$, p<.01, ω^2 =.09. Tukey HSD tests indicated that participants in the No-KR control condition (M= -1.25) reported a greater decline in *Task Engagement* after the vigil than those in the composite-KR condition (M= -.38; Cohen's d=1.02). No other comparisons reached conventional levels of significance (p> .05 in each case).

In sum, only Hit- and Composite-KR were effective in enhancing perceptual sensitivity relative to a no-KR control, and only Composite-KR was effective in reducing the stress associated with the vigil. However, this effect was restricted to the *Task Engagement* dimension. In subsequent sections the effects of pessimism and optimism are evaluated to determine whether they moderated these relations between KR, performance, and stress state.

Pessimism and Optimism

The mean pessimism score across all 110 participants was 36.5 (SD=6.6), and the mean optimism score was 53.56 (SD=5.0). The correlation between these variables was r = -.40 (p<.001), which is lower than that usually obtained between these two variables (approximately r =.5; see Dember et al., 1989; Dember, 2002). However, inspection of the mean, standard deviation, and coefficient α for each scale (coefficient α=.81 for optimism; α=.83 for pessimism) indicated that these indices did not meaningfully differ from results obtained from prior samples. Preliminary analyses indicated that there were no significant differences among KR conditions in optimism or pessimism (p>.5 in each case).

Feedback and Dispositional Optimism and Pessimism

To evaluate the impact of Optimism and Pessimism, and their interaction with KR format, a hierarchical regression approach was employed using procedures outlined by Pedhazur (1997; see also Hardy, 1993). The KR variable was dummy-coded so that the No-KR condition was assigned a value of zero across all four vectors. Each vector (i.e., Hit-KR, False Alarm-KR, Miss-KR, and Composite-KR) therefore represented the comparison between a KR condition and the No-KR control. The terms in each regression were trait (optimism or pessimism), KR (the dummy coded vectors), and trait by KR product vector terms ('interaction' terms). Separate regressions were computed with optimism and pessimism to simplify analyses and avoid problems associated with multi-collinearity.

When the vector products are included in the model, interpretation of the regression weights for KR and trait terms are meaningless, since the product vectors are correlated with

their constituent parts (see Pedhazur, 1997; see also Jaccard & Turrisi, 2003). Thus, the interaction terms are interpreted according to the increment in the proportion of variance accounted for (i.e., ΔR^2) after the variance accounted for by pessimism and the KR vectors (the 'main effects') has been entered into the model. Step 1 of the hierarchical regression procedure consisted of entry of the trait variable (optimism or pessimism), step 2 consisted of entry of the trait variable and the dummy coded KR variable, and step 3 consisted of the main effects and the 'interaction' terms. A non-significant increment in ΔR^2 for the full model (step 3) resulted in evaluating the models at step 1 (trait only) or step 2 (trait and KR). Cases in which a significant product vector was obtained indicated that the slope of the regression of the dependent variable on the trait variable was significantly different between the corresponding KR conditions. That is, the test for the regression weight for each product vector (e.g., Hit-KR by pessimism, False Alarm-KR by pessimism) constitutes a test of the difference between the regression weights for the No-KR group and the particular KR condition. Following Pedhazur (1997), such product vectors were tested using the Johnson-Neyman procedure for determination of simultaneous regions of significance (see also Huitema, 1980; Jaccard & Turrisi, 2003). This procedure, analogous to a test of simple effects in analysis of variance, involves determining over which regions of the continuous predictor variable (e.g., pessimism) differences between the categorical variables (e.g. No-KR versus FA-KR) are significant. In this framework, the trait variable is considered a moderating variable for the relation between KR and the dependent variable (i.e., performance, stress; see Jaccard & Turrisi, 2003). To control for Type I error, the Bonferroni F statistic was used in the computation of regions of significance (see Huitema, 1980). To simplify performance analyses, regressions of performance on the trait variable (pessimism, optimism), KR, and the associated product vectors were performed for sensitivity and response bias in two

ways: 1) an overall index of sensitivity and bias based on the total number of correct detections and false alarms over the entire vigil; and 2) a difference score for these indices between the first and last periods on watch.

Pessimism

Performance

Regressions of overall A' and the difference score between periods 1 and 6 revealed no impact of pessimism on performance, nor did pessimism interact with KR in influencing performance (p>.05 in each case). Similar analyses with overall β_d" and the difference score for periods 1 and 6 indicated that pessimism did not impact response bias, nor did it interact with KR to influence overall response bias or change in bias between the beginning and end of the session (p>.05 in each case).

Perceived Stress

Stress state was analyzed via separate hierarchical regressions of the pre-post difference scores of each scale (*Distress*, *Task Engagement*, and *Worry*) on pessimism, KR group, and the interaction between these factors.

Distress. The regression of pre-post *Distress* on pessimism, KR, and the product vectors (i.e., KR x pessimism) was significant, $F(9, 100)=2.41$, $p<.05$, $R^2=.18$. The increment in variance accounted for by the product vector terms was significant, $F(4, 100)=2.59$, $p<.05$ $\Delta R^2 = .09$. Significant regression weights were observed for the dummy coded false alarm vector (b=-5.35, $SE_b=2.06$, β=-2.00, t(34)= -2.00, $p<.05$) and the product vector for pessimism and False Alarm-KR (b=.14, $SE_b=.06$, β=1.92, t(34)= 2.46, $p<.05$). The significant effect for the regression weight

associated with the product vector for pessimism and False Alarm-KR indicates that the regression weights for the No-KR and False Alarm-KR conditions were significantly different from one another. The relation between pre-post task *Distress* and pessimism is shown in Figure 4 for the False Alarm-KR and No-KR conditions. Application of the Johnson-Neyman procedure revealed that for pessimism scores less than 31.7 observers who received False Alarm-KR reported significantly less *Distress* than those in the No-KR group. Hence, False Alarm-KR was effective in reducing *Distress*, but only for participants low in pessimism. This effect was attenuated for individuals high in pessimism.

Insert Figure 4 about here

Task Engagement. The regression of pre-post *Task Engagement* on pessimism, KR, and the product vectors was significant, $F(9, 100)=3.57$, $p<.01$, $R^2= .24$. The increment in variance accounted for by the product vector terms was significant, $F(4, 100)=3.33$, $p<.05$ $\Delta R^2 = .11$. Significant regression weights were observed for the dummy coded false alarm vector (b=4.64, $SE_b=1.64$, $\beta=2.08$, $t(34)= 2.82$, $p<.01$) and the product vector for pessimism and false alarm KR (b= -.11, $SE_b=.04$, $\beta= -1.85$, $t(34)= -2.47$, $p<.05$). Pre-post *Task Engagement* z-scores are plotted as a function of pessimism for the False Alarm-KR and No-KR conditions in Figure 5. The Johnson-Neyman procedure was applied to investigate the significant False Alarm-KR/pessimism product vector. This analysis revealed that for pessimism scores below 35.6 observers who received False Alarm-KR reported significantly higher *Task Engagement* than observers who did not receive feedback. That is, False Alarm-KR was more effective in increasing *Task Engagement* for participants relatively low in pessimism, but the effectiveness of this KR-format was reduced for individuals higher in pessimism.

Insert Figure 5 about here

Worry. Analyses indicated non-significant regressions for pessimism and KR (p>.05 in each case), but a significant regression for the full model (step 3), $F(9, 100)=2.23$, $p<.05$, $R^2=.18$. The increment in variance accounted for by the product vector terms was significant, $F(4, 100)=3.79$, $p<.01$ $\Delta R^2=.13$. This outcome is analogous to non-significant main effects in an ANOVA but a significant interaction. However, tests of the regression weights revealed no significant b's, although the coefficient for the product vector for Miss-KR/pessimism was marginally significant (b=-.08, SE_b=.05, β=-1.41, t(34)=-1.87, p=.06). Pre-post Worry scores are plotted as a function of pessimism for the miss KR and no-KR conditions in Figure 6. Application of the Johnson-Neyman procedure was unsuccessful, likely due to high within-groups variability, which can cause failures in solutions for regions of significance (see Pedhazur, 1997). Separate regressions of *Worry* on pessimism for the No-KR and Miss-KR conditions revealed a non-significant regression for the former and a significant regression for the latter, $F(1, 20)=9.11$, $p<.01$, $R^2=.31$ (b=-.06, SE_b=.02, β=-56, t(34)=-3.09, p<.01). Contrary to prediction, increased pessimism was related to decreased *Worry* for participants in the Miss-KR condition.

Insert Figure 6 about here

Optimism

Performance

Regressions of overall A' on optimism, KR, and their product vectors revealed a significant regression for the full model, $F(9, 100)= 2.73$, $p<.01$. However, inspection of the data

indicated that these effects were primarily due to two outlier scores in the False Alarm-KR condition. When these were omitted from the analysis, no significant effects were observed. Inspection of the DSSQ and OPI scores of these two individuals indicated that they did not differ from their group mean (the False Alarm-KR condition) on the OPI or DSSQ scales by more than 1.5 standard deviations. Similar analyses on A' difference scores, β_d'', and β_d'' difference scores indicated that optimism did not significantly influence change in performance between the beginning and end of the vigil, nor did optimism interact with KR format in influencing overall response bias or changes or response bias (p>.05 in each case).

Perceived Stress

As with the analyses involving pessimism, stress state was analyzed via separate hierarchical regressions of each scale (*Distress*, *Task Engagement*, and *Worry*) on optimism, KR, and the interaction between these factors.

Distress and Task Engagement. Regression analysis indicated that optimism, KR, and the product vectors did not significantly predict *Distress* (p>.05). Optimism predicted pre-post *Task Engagement*, $F(1,108)=6.50$, $p<.05$, $R^2=.06$, and the model including optimism and KR conditions was significant, $F(5, 104)=3.84$, $p<.05$, $R^2=.16$, although the product terms for these variables did not significantly add to the variance accounted for by the regression (p>.05). Significant regression weights were observed for optimism, (b= -.03, SE_b=.02, β=-.19, t(38)= -2.08, p<.05), the False Alarm-KR vector (b=.56, SE_b=.26, β=.25, t(38)= 2.18, p<.05), and the Composite-KR vector (b=.79, SE_b=.26, β=.36, t(38)= 3.09, p<.01).

Worry. For pre-post task *Worry*, the full regression model was significant, $F(9, 100)=2.62$, $p<.01$, $R^2=.19$, and the product vectors significantly added to the variance accounted

for by the regression, $F(4, 100)=4.83$, $p<.01$, $\Delta R^2=.16$. Significant regression weights were observed for the False Alarm-KR vector (b=7.39, $SE_b=2.72$, β=3.26, t(34)= 2.72, $p<.01$), and the False Alarm-KR/optimism product vector (b= -.14, $SE_b=.05$, β= -3.34, t(34)= -2.84, $p<.01$). Pre-post *Worry* scores associated with the False Alarm-KR and No-KR conditions are plotted as a function of optimism in Figure 7.

Insert Figure 7 about here

Johnson-Neyman analyses revealed that for optimism scores below 28.6 False Alarm-KR induced greater *Worry* compared to No-KR, but for scores above 57.6 receiving False Alarm KR induced less worry relative to No-KR controls. Note that the results for low end of the optimism scale must be interpreted with caution since the regions of significance represent an extrapolation to a region in which there are no sample data.

DISCUSSION

A primary purpose for the present study was to investigate whether the performance differences as a function of KR format observed in previous research (Dittmar et al., 1985) were associated with differences in stress state, and whether combining these formats into a single condition is most effective in improving performance and reducing stress. A second goal was to investigate the possibility, based upon hypotheses derived from FIT (Kluger & DeNisi, 1996), that dispositional optimism and pessimism would influence the effectiveness of different KR formats in improving vigilance performance and in reducing the stress associated with prolonged monitoring. Specifically, it was hypothesized that False Alarm- and Miss-KR should impair

performance and increase stress for individuals high in pessimism. By contrast, Hit-KR should facilitate performance and reduce stress symptoms for individuals high in optimism, although these effects were expected to be weaker than those associated with pessimism.

The present results obtained with Hit-KR were consistent with the findings of Dittmar et al. (1985), while the Composite-KR was effective at enhancing sensitivity as also observed in prior testing (Szalma et al., 1998; Szalma et al., 2004). Also consistent with Dittmar et al. (1985) Miss-KR was not effective in enhancing perceptual sensitivity. However, in contrast to the results of Dittmar and her colleagues, False Alarm-KR was also not effective in enhancing sensitivity relative to a no-KR control. Moreover, a major finding of Dittmar and her colleagues (1985) of an interaction between KR format and time on task for both sensitivity and response bias was not confirmed in the present experiment. This is most probably due to the absence here of a sensitivity decrement and large within-groups variability observed in the response bias data. These differences are likely task specific and confirm that the influence of KR format depends on the information processing requirements of the particular vigilance task at hand. The task employed by Dittmar et al. (1985) required discrimination of the length of a vertical line, while in this experiment participants monitored patterns of change in the shape of an object display. Prior research has shown that use of an object display enhances sensitivity and attenuates the vigilance decrement without the provision of feedback (Szalma, 2002). It might be that the differential effects of KR-format depend on the perceptual and cognitive demands of the specific task and while several of these factors have been determined (Davies & Parasuraman, 1982; Warm & Dember, 1998) a full exposition has yet to emerge. In comparison with the Dittmar et al. (1985) results, note that the mode of KR also has varied since their report. Dittmar and her colleagues used a 1000 Hz tone as feedback, while in the present experiment a female voice

controlled by the computer was employed to deliver feedback. Whether such mode differences in KR presentation influences effectiveness for enhancing vigilance is a matter for future research.

Stress

Consistent with previous vigilance experiments, observers in this study found the vigil to be stressful (Hancock & Warm, 1989). A significant increase in *Distress* and reduction in *Task Engagement* was observed across all conditions. However, KR-format failed to significantly impact *Distress* or *Worry*. A significant KR effect was observed for *Task Engagement*, but post-hoc tests revealed that this effect was restricted to the Composite-KR condition. Taken together with the performance results, these results indicate that although Hit-, Miss-, and False Alarm-KR differ in their effects, combining them has the greatest impact in improving performance and reducing stress. Moreover, these results indicate that regardless of the separate effects of the different KR-formats, training regimens for vigilance may benefit for providing feedback regarding all three outcomes. Providing such information may direct attention away from the self and toward task-goals to a greater degree than when the formats are presented singly. The different forms of KR in this instance might serve as cues for observers to attend to their respective responses to stimuli, so that the KR serves to strengthen the association between stimulus characteristics and observer response. The different effects across the three stress state dimensions reflect the multidimensional nature of the stress construct (Matthews, 2001), and may represent different appraisals to task demands and different coping responses to those demands. Thus, the 'core relational theme' (Lazarus, 1991) that links to *Distress* is 'overload of capacity for successful performance' and *Task Engagement* relates to 'commitment of effort' (Matthews et al., 1999; Matthews et al., 2002). Thus, composite-KR did not relieve the overload

associated with vigilance (see Warm, Dember, & Hancock, 1996), but it did enhance the commitment of effort to perform. The latter result conforms to the exposition concerning the motivational effect that KR has on vigilance (see Warm & Jerison, 1984).

In this experiment, Worry declined after the vigil across all conditions. According to Matthews et al. (1999) the *Worry* dimension is associated with self evaluation, 'such that the nature of the task provokes re-assessment of personal qualities and goals.' (p. 348) and with emotion-focus and avoidant coping. Indeed, they noted that self-reflection is most likely to occur in monotonous tasks such as vigilance. Further, Matthews et al. (2002) argued that this theme would fall under the 'meta task' process described by Kluger and DeNisi (1996). In this experiment Worry declined across all conditions, indicating an overall attenuation of 'meta task' processes. However, these effects depended on the pessimism and optimism of observers, as discussed below.

Performance and Stress: The Non-Reciprocity of Hit- and False Alarm-KR

Dittmar and her colleagues noted that Hit- and False Alarm-KR are reciprocals of one another since they provide the same information regarding whether a response was correct. However, their effects in this study were not equivalent in terms of performance and self-report of stress state. Hit-KR, but not False Alarm-KR, enhanced sensitivity relative to observers who did not receive KR. False Alarm-KR, but not Hit-KR, reduced the *Distress* and increased *Task Engagement* relative to no-KR controls, but only for individuals low in pessimism. Although Hit-KR observers performed better than controls who did not receive KR, this performance enhancement was not associated with reduced *Distress* or increased *Task Engagement*, suggesting that Hit-KR does not alter the coping strategies observers use to deal with task

demands. In contrast, False Alarm-KR reduced the stress symptoms for individuals low in pessimism, but this benefit was not associated with performance effects. This indicates that False Alarm feedback influences appraisals observers make in respond to task demands, that these appraisals occur independent of actual performance, and that these appraisals depend upon the person's level of dispositional pessimism.

Effects of Pessimism and Optimism: The Test of Feedback Intervention Theory (FIT)

Performance

Consistent with prior experimental findings, Pessimism and Optimism did not significantly impact the performance of observers (see Ganey et al., 2003; Szalma, 2002; Thropp et al., 2003). In one exception, Helton et al. (1999) observed a steeper decrement in pessimists relative to optimists, but evidence from this and other investigations imply that these effects may be dependent on task characteristics. Specifically, Helton et al. (1999) employed a rather difficult discrimination at a very high event rate (57.5/min), while relatively easier discriminations at lower event rates were used in this and other experimental procedures which did not show this influence.

Helton and his colleagues also observed that pessimism/optimism did not interact with false summary KR to influence performance, a finding that was replicated in this study using veridical feedback after each response (or failure to respond appropriately). This result was inconsistent with the hypotheses based on FIT. One possible explanation is that the nature of the KR and the task was not sufficiently important relative to the individuals' self-goals (i.e., *meta-task* goals). This interpretation is supported by the finding that across all conditions pre-post

Worry declined, which reflects less self-evaluative processes (Matthews et al., 2002). Stronger effects might be observed if the attention to the levels of the goal hierarchy discussed by Kluger and DeNisi (1996; i.e., *meta-task*, *task-motivational*, and *task-learning processes*) were manipulated specifically.

A second explanation can be derived from the processing efficiency theory of Eysenck and Calvo (1992). According to this theory, individuals who have difficulty coping with stress can compensate by exerting more task effort. The effects of stress are therefore manifested as a reduction in the efficiency of information processing rather than in the output of such processing. Thus, two individuals may exhibit equivalent levels of performance, but one individual is maintaining that performance at greater cost. Evidence from the current study indicates that individuals high in pessimism can achieve levels of performance equivalent to those of their less pessimistic cohorts but that maintaining such performance occurs at greater cost, particularly when negative feedback is provided. It may therefore be that negative KR and pessimism interact to influence processing efficiency rather than the absolute level of performance.

Stress

Pessimism. Although the expected interaction between trait and KR was not observed in regard to performance, interactions were observed in stress state and the effect of pessimism on stress depended on the joint effects of KR format and dimension of stress. Thus, the effect of 'negative KR' (i.e., False Alarm- and Miss-KR) in reducing stress was observed only for individuals low in pessimism. Moreover, False Alarm-KR exerted its influence on *Distress* and *Task Engagement*, while Miss-KR exerted its influence on *Worry*. However, while False Alarm –KR reduced *Distress* and increased *Task Engagement* for individuals low in pessimism, Miss-

KR reduced *Worry* for individuals *higher* in that trait. Thus, the two forms of negative KR influenced different stress state dimensions in different ways. It is important to remember that *Distress* and *Task Engagement* dimensions relate to mood, motivation, and cognitive states, but the *Worry* dimension reflects only the cognitive component of affective states (Matthews et al., 1999; Matthews et al., 2002). False Alarm-KR may therefore have pervasive effects across the mood, motivation, and cognitive components of state, while Miss-KR may influence primarily the cognitive component. The latter result does not conform to the original hypotheses, based on the findings of Dittmar et al. (1985) and Kluger and DeNisi (1996), that one would expect Miss-KR would cause participants to engage in more of the '*meta task*' processes associated with the *Worry* dimension (Matthews et al., 2002), and that these effects would be more likely for those higher in pessimism. However, individuals higher in pessimism who received Miss-KR reported *less Worry* after the vigil, indicating that they were engaging in less self-evaluative processes.

The interaction of KR with pessimism may also be influenced by the different coping strategies associated with the three dimensions of stress state. *Distress* and the absence of *Task Engagement* are associated with emotion-focused coping (Matthews et al., 1999), and this form of coping is also typically adopted by pessimists (Helton et al., 1999; Scheier & Carver, 1987). Further, Szalma (2002) indicated that pessimism is associated with increased reliance on emotion-focused coping strategies. Thus, the differences between False Alarm-KR and no-KR on the *Distress* and *Task Engagement* dimensions may be due to less emotion focused coping among those low in pessimism. As pessimism increases, emotion focused coping increases, diverting processing resources to *meta-task processes* (Kluger & DeNisi, 1996), and thereby reduces the effectiveness of KR in reducing these symptoms.

Optimism. As expected, the interaction effects between KR and optimism were weaker than those between KR and pessimism. For participants who received False Alarm-KR, higher levels of optimism were associated with decreased *Worry* and lower levels of optimism associated with increased *Worry*. However, higher optimism predicted lower *Task-Engagement* across all KR conditions. Taken together, these results indicate that individuals high in optimism find False-Alarm KR less aversive than those low on that trait, but optimism does not protect against loss of Engagement. The reasons for the latter effect are a matter for future research.

Summary and Conclusions

It is evident that the stress effects of vigilance are complex and multidimensional and that the effectiveness of KR in reducing stress depends on both the type of KR provided and characteristics of the individual. Taken together with the results of prior research in this area, the current study indicates that these KR effects depend on the perceptual demands of the task. The results of the present study again serve to underscore the importance of examining individual differences in stress responses to vigilance (Berch & Kanter, 1984). KR format did not significantly impact *Distress* or *Worry* dimensions of stress, but KR effects on these variables did emerge when dispositional pessimism and optimism were taken into account. Further investigation needs to be directed to explore these joint effects of feedback and observer characteristics on performance and stress in vigilance tasks. Such efforts can lead to improved training regimens that are tailored to the individual to maximize performance in crucial operational settings.

The hypotheses based upon Feedback Intervention Theory received partial support, in that negative feedback did not enhance performance, and that False Alarm-KR was effective in

reducing stress symptoms only for those individuals low in pessimism. However, the predicted interaction effect on performance of feedback type and pessimism was not observed. Future investigations should test for this interaction by systematically manipulating participants' goals according to the hierarchy described in Kluger and DeNisi (1996).

Since 9/11, the world has become much more wary and invested in safety and security systems. Often, these individuals and the technology created to help them perform their tasks are looking for the equivalent of a 'needle in a haystack.' Even with highly sensitive systems and highly motivated observers, such rare signal occurrences lead to an abundance of false alarms (Parasuraman, Hancock, & Olofinboba, 1997) that eventually society in general will not tolerate. From our present experiment it is clear that the effectiveness and efficiency of vigilance is contingent upon the character of the targets being searched, how that information is presented, and the nature of the individual to whom it is presented. The current findings point to the strong, interactive nature of these factors and begin to chart a course to better systems, better selection, and better training for those who perform this vital task.

ACKNOWLEDGEMENT

Completion of this work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. P.A. Hancock, *Principal Investigator.* The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

REFERENCES

Bennett, K.B., & Flach, J.M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34*, 513-533.

Bennett, K.B., Nagy, A.L., & Flach, J.M. (1997). Visual displays. In: G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 659-696). New York: Wiley.

Berch, D.B., & Kanter, D.R. (1984). Individual differences. In: J.S. Warm (Ed.), *Sustained attention in human performance.* (pp. 143-178). Chichester: Wiley.

Carver, C.S., & Scheier, M.F. (1981). *Attention and self-regulation: A control theory approach to human behavior*. New York: Springer-Verlag.

Carver, C.S., & Scheier, M.F. (1998). *On the self-regulation of behavior*. New York: Cambridge University Press.

Chang, E.C. (2002, Ed.). *Optimism and pessimism: Implications for theory, research, and practice*. Washington, DC: American Psychological Association.

Craig, A. (1977). Broadbent and Gregory revisited: Vigilance and statistical decision. *Human Factors, 19*, 25-36.

Craig, A. (1978). Is the vigilance decrement simply a response toward probability matching? *Human Factors, 20,* 441-446.

Craig, A. (1979). Nonparametric measures of sensory efficiency for sustained monitoring tasks. *Human Factors, 21*, 69-78.

Davies, D.R., & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Dember, W.N. (2002). The optimism-pessimism instrument: Personal and social correlates. In: E.C. Chang (Ed.), *Optimism and pessimism: Implications for theory, research, and practice.* (pp. 281-299). Washington, DC: American Psychological Association.

Dember, W.N., Martin, S.H., Hummer, M.K., Howe, S.R., & Melton, R.S. (1989). The

measurement of optimism and pessimism. *Current Psychology: Research and Reviews, 8,* 102-119.

Dittmar, M.L., Warm, J.S., & Dember, W.N. (1985). Effects of knowledge of results on performance in successive and simultaneous vigilance tasks: A signal detection analysis. In: R.E. Eberts and C.G. Eberts (Eds.), *Trends in ergonomics/human factors II.* (pp. 195-202). Amsterdam: Elsevier.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121*, 275-277.

Eysenck, M.W., & Calvo, M.G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*, 409-434.

Ganey, H.C.N., Szalma, J.L., Hancock, P.A., Mouloua, M., Connolly, A., Dalton, J., & Davis, F. (2003). The role of choice and individual differences in the mitigation of noise and task stressors in vigilance. *Proceedings of the Human Factors and Ergonomics Society*, *47*, 1078-1081.

Hancock, P.A., & Hart, S.G. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in Design, 10,* 6-16.

Hancock, P.A., & Szalma, J.L. (2003). Operator stress and display design. *Ergonomics in Design*, *11*, 13-18.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*, 519-537.

Hardy, M.A. (1993). *Regression with dummy variables*. Newbury Park, CA: Sage.

Harris, D. (2002). How to really improve airport security. *Ergonomics in Design*, *10*, 17-22.

Helton, W.S., Dember, W.N., Warm, J.S., & Matthews, G. (1999). Optimism-pessimism and

false failure feedback: Effects on vigilance performance. *Current Psychology, 18*, 311-325.

Huitema, B.E. (1980). *The analysis of covariance and alternatives*. New York: Wiley.

Hummer, M.K., Dember, W.N., Melton, R.S., & Schefft, B.K. (1992). On the partial independence of optimism and pessimism. *Current Psychology: Research & Reviews, 11*, 37-50.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression, 2nd Ed*. Thousand Oaks, CA: Sage.

Kluger, A.N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.

Lazarus, R.S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lazarus, R.S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer-Verlag.

Macmillan, N.A., & Creelman, C.D. (2005). *Detection theory: A user's guide*. (2nd Edition). Mahwah, NJ: Erlbaum.

Matthews, G. (2001). Levels of transaction: A cognitive science framework for operator stress. In: P.A. Hancock and P.A. Desmond (Eds.), *Stress, workload, and fatigue.* (pp. 5-33). Mahwah, NJ: Erlbaum.

Matthews, G., & Campbell, S.E. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society 42*, 821-825.

Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., & Warm, J.S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, *2*, 315-340.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999).

Validation of a comprehensive stress state questionnaire: Towards a state 'big three'? In: I. Mervielde, I.J. Deary, F. DeFruyt, and F. Ostendorf (Eds.), *Personality psychology in Europe (vol 7)* (pp. 335-350). Tilburg: Tilburg University Press.

Maxwell, S.E., & Delaney, H.D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd edition). Mahwah, NJ: Erlbaum.

Parasuraman, R., Hancock, P.A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision warning systems. *Ergonomics, 39*, 390-399.

Pedhazur, E.J. (1997). *Multiple regression in behavioral research*, 3rd ed. Ft. Worth: Harcourt Brace.

See, J.E., Howe, S.R., Warm, J.S., & Dember, W.N. (1995). A meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117*, 230-249.

See, J.E., Warm, J.S., Dember, W.N., & Howe, S.R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors, 39*, 14-29.

Scheier, M.F., & Carver, C.S. (1987). Dispositional optimism and physical well being: the influence of generalized outcome expectancies on health. *Journal of Personality, 55*, 170-210.

Szalma, J.L. (2002a). Individual differences in the stress and workload of sustained attention. *Proceedings of the Human Factors and Ergonomics Society*, *46*, 1002-1006.

Szalma, J.L. (2002b). Workload and stress of configural displays in vigilance tasks. *Proceedings of the Human Factors and Ergonomics Society*, *46*, 1536-1540.

Szalma, J.L., Hancock, P.A., Warm, J.S., Dember, W.N., & Parsons, K.S. (2005). *Training for vigilance: Using predictive power to evaluate feedback effectiveness*. Manuscript submitted to *Human Factors*, March, 2005.

Szalma, J.L., Miller, L.C., Hitchcock, E.M., Warm, J.S., & Dember, W.N. (1998). Intraclass and

interclass transfer of training for vigilance. In: M.W. Scerbo and M. Mouloua (Eds.), *Automation Technology and Human Performance.* (pp. 183-187). Mahwah, NJ: Erlbaum.

Szalma, J.L., Warm, J.S., Matthews, G., Dember, W.N., Weiler, E.M., Meier, A., & Eggemeier, F.T. (2004). Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Human Factors, 46,* 219-233.

Temple, J.G., Warm, J.S., Dember, W.N., Jones, K.S., LaGrange, C.M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors, 42,* 183-194.

Thropp, J.E., Szalma, J.L., Ross, J.M., & Hancock, P.A. (2003). Individual differences in dispositional pessimism, stress, and coping as a function of task type. *Proceedings of the Human Factors and Ergonomics Society*, *47*, 1073-1077.

Warm, J.S. (1993). Vigilance and target detection. In: B.M. Huey and C.D. Wickens, (Eds.), *Workload transition: Implications for individual and team performance.* (pp. 139-170). Washington, DC: National Academy Press.

Warm, J.S., & Dember, W.N. (1998). Tests of vigilance taxonomy. In: R.R. Hoffman, M.F. Sherrick, and J.S. Warm (Eds.). *Viewing psychology as a whole: The integrative science of William N. Dember.* (pp. 87-112). Washington, DC: American Psychological Association.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In: R. Parasuraman and M. Mouloua (Eds.), *Automation and human performance*. (pp. 183-200). Mahwah, NJ: Lawrence Erlbaum.

Warm, J.S., & Jerison, H.J. (1984). The psychophysics of vigilance. In: J.S. Warm (Ed.),

Sustained attention in human performance. (pp. 15-59). Chichester, United Kingdom: Wiley.

Figure Captions

Figure 1. Stimuli employed in the current experiment. Displays (a) through (e) are 'neutral events,' requiring no overt response from the observer. Displays (f) and (g) are 'critical signals,' defined by simultaneous changes on both axes. *Note*. P = Pressure; T = Temperature; FI = Flow Rate In; FO = Flow Rate Out.

Figure 2. Perceptual sensitivity (A') as a function of period on watch. *Note*. NKR = No KR.

Figure 3. Response bias (β_d") as a function of period on watch. *Note*. NKR = No KR.

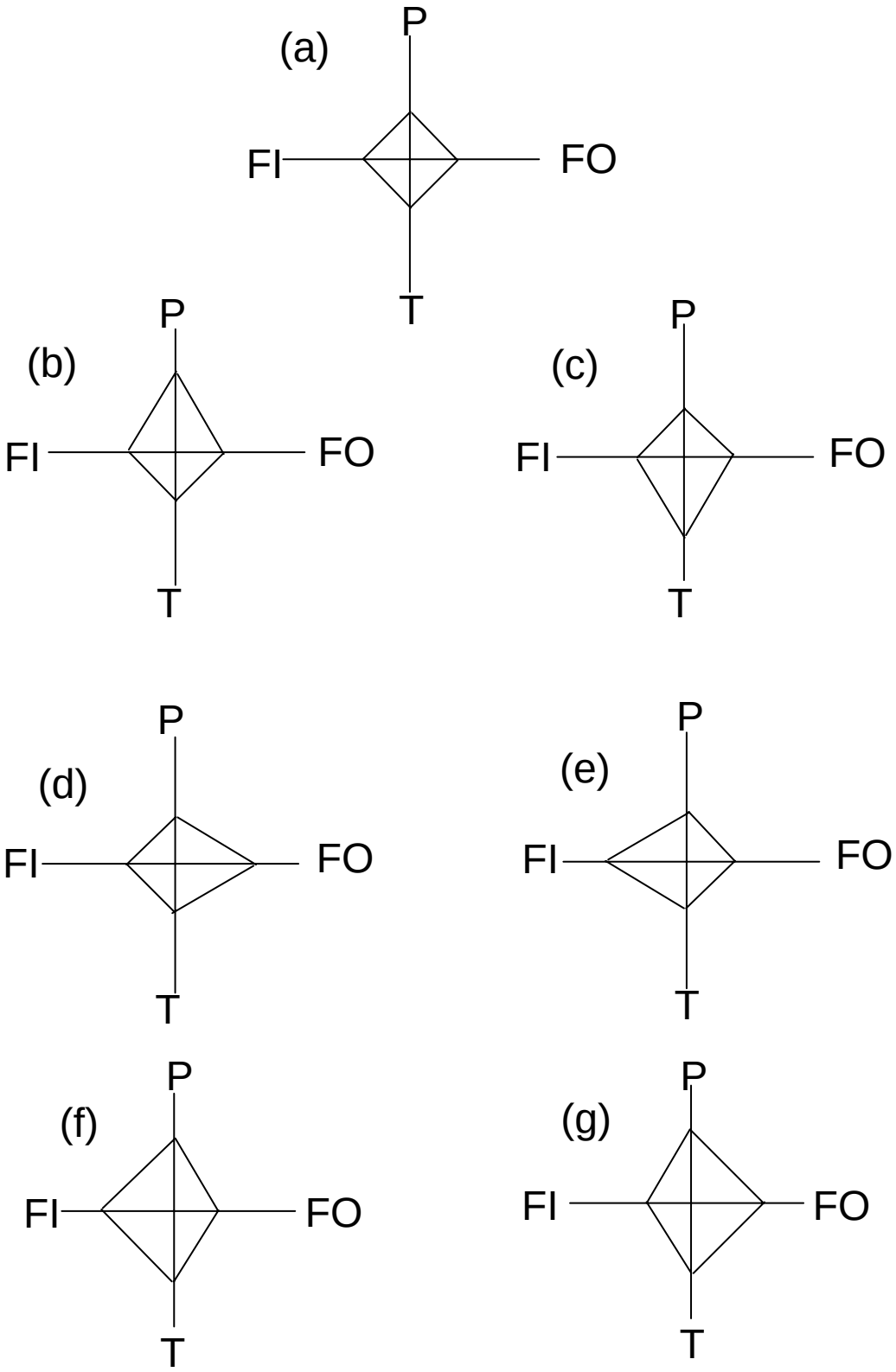
Figure 4. Pre-Post task *Distress* as a function of pessimism for the False Alarm-KR and No-KR groups. *Note*. Dotted line with arrow represents the region of significant differences between the regression lines. Separate regression equations: False Alarm-KR group: D=.06P-1.64, R^2=.17; No-KR group: D= -.08P+3.71, R^2=.13; Intersection point of regression lines: P=38.8. NKR = No-KR; FAKR = False Alarm KR; D = Distress; P = Pessimism.

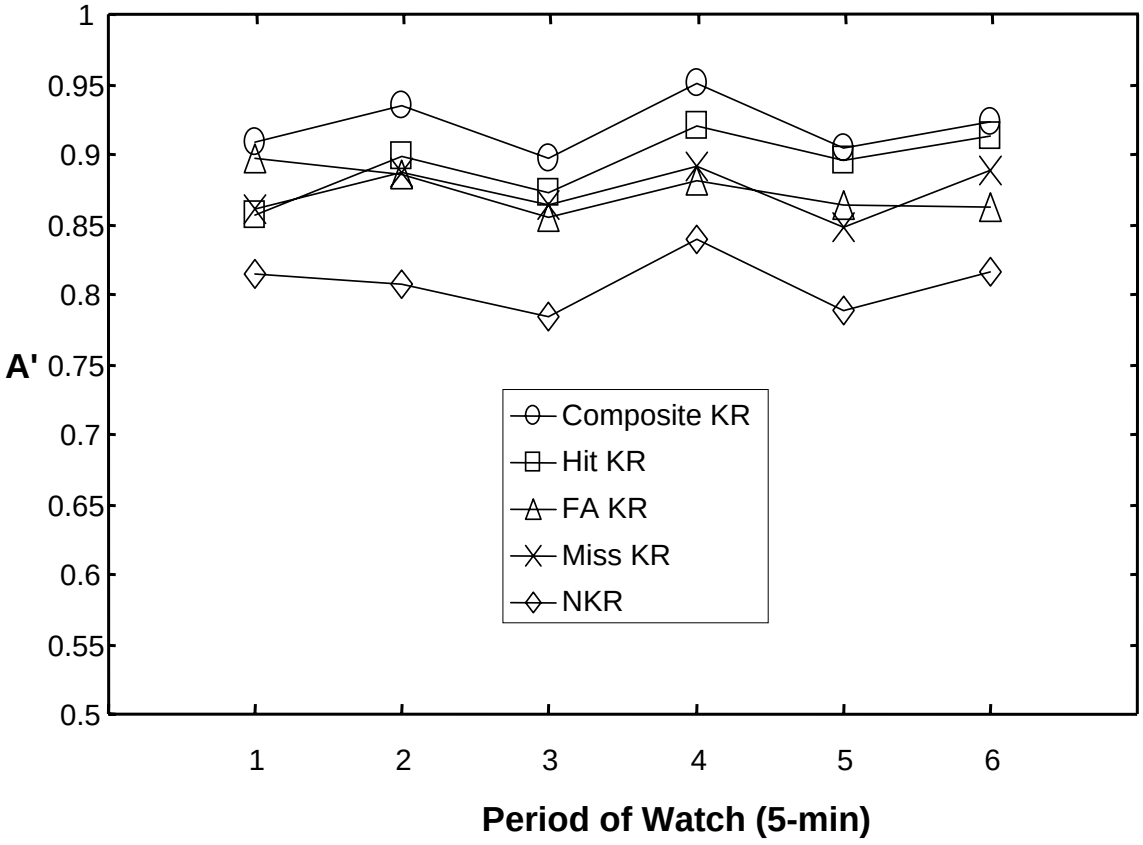
Figure 5. Pre-Post *Task Engagement* as a function of pessimism for the False Alarm-KR and No-KR groups. *Note*. Dotted line with arrow represents the region of significant differences between the regression lines. Separate regression equations: False Alarm-KR group: TE= -.05P+1.32, R^2=.14; No-KR group: TE= .06P-3.32, R^2=.11; Intersection point of regression lines: P=41.8. NKR = No-KR; FAKR = False Alarm KR; TE = Task Engagement; P = Pessimism.

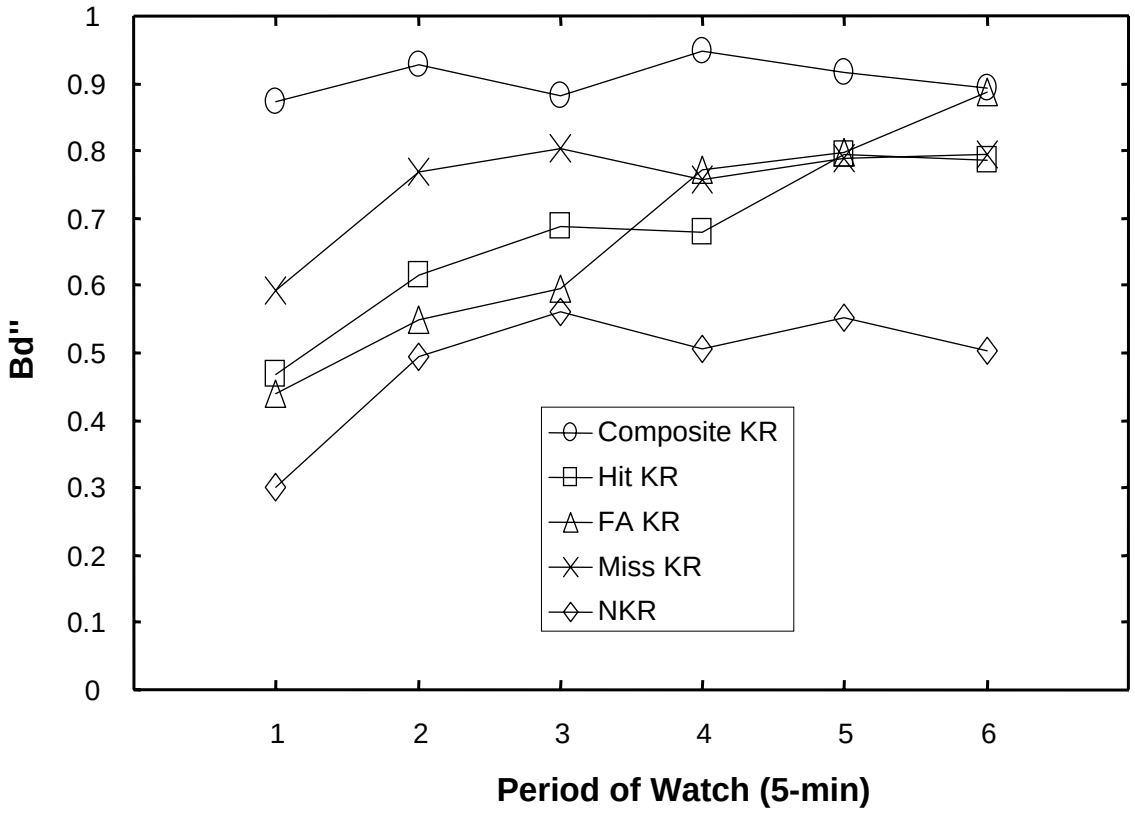
Figure 6. Pre-Post task *Worry* as a function of pessimism for the Miss-KR and No-KR groups. *Note*. Separate regression equations: Miss-KR group: W= -.06P+1.76, R^2=.31; No-KR group: W= .02P - .96, R^2=.01; Intersection point of regression lines: P=32.3. NKR = No-KR; W = Worry; P = Pessimism.

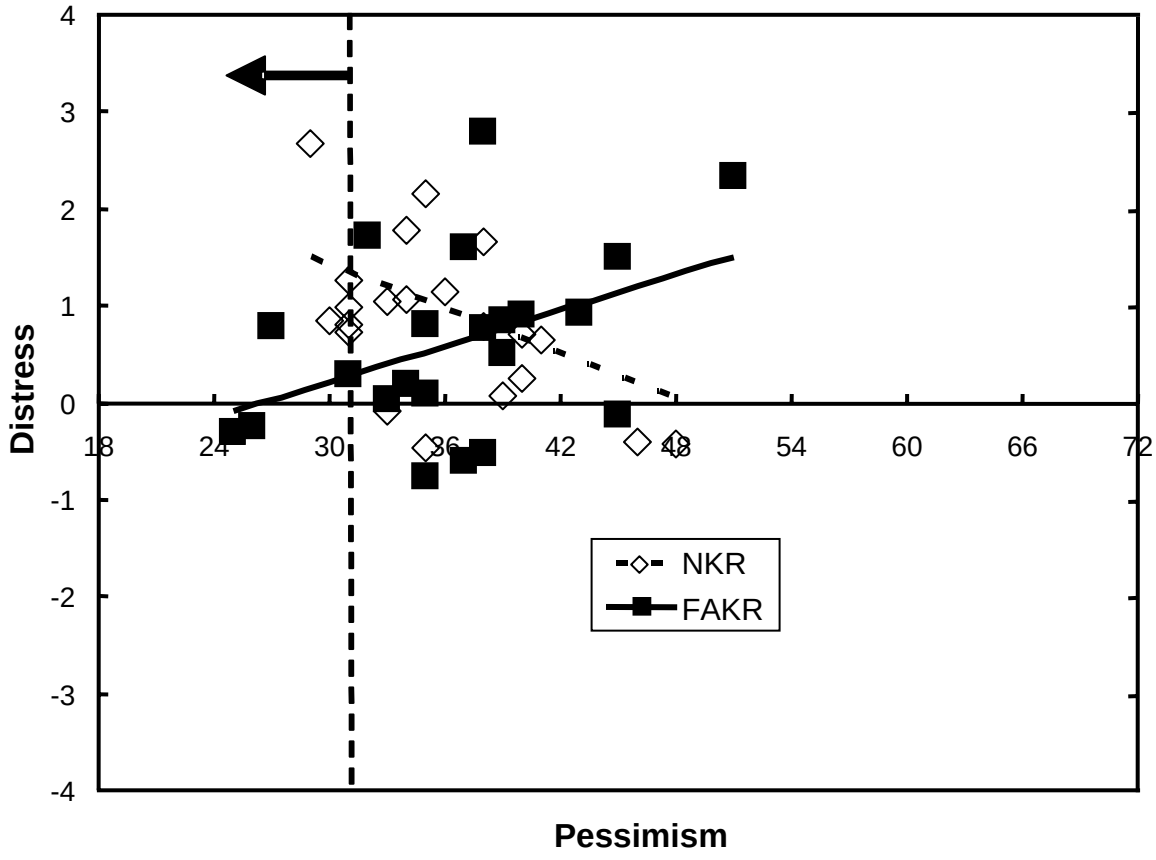
Figure 7. Pre-Post task *Worry* as a function of optimism for the False Alarm-KR and No-KR groups. *Note*. Dotted lines with arrows represent the simultaneous regions of significant

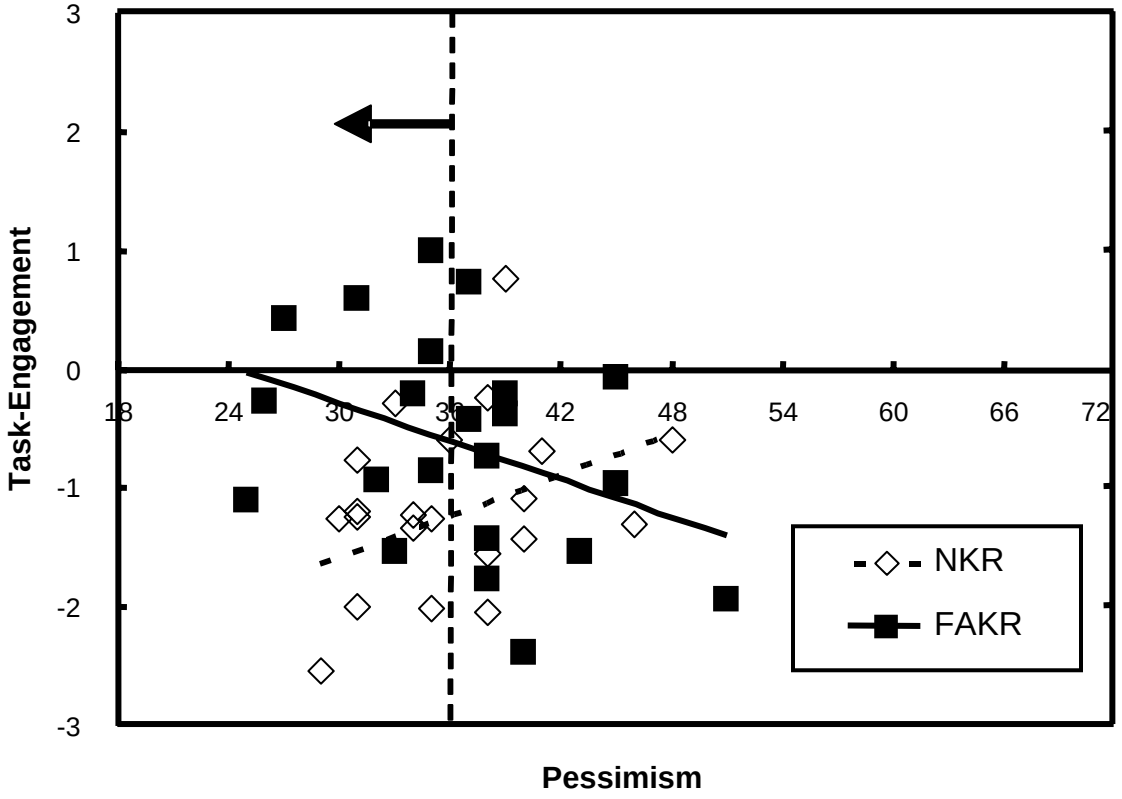
differences between the regression lines. Separate regression equations: False Alarm-KR group: W=-.11Opt+5.26, R^2=.41; No-KR group: W= .04Opt-2.12, R^2=.04; Intersection point of regression lines: Opt=51.8. NKR = No-KR; FAKR = False Alarm KR; W = Worry; Opt = Optimism.

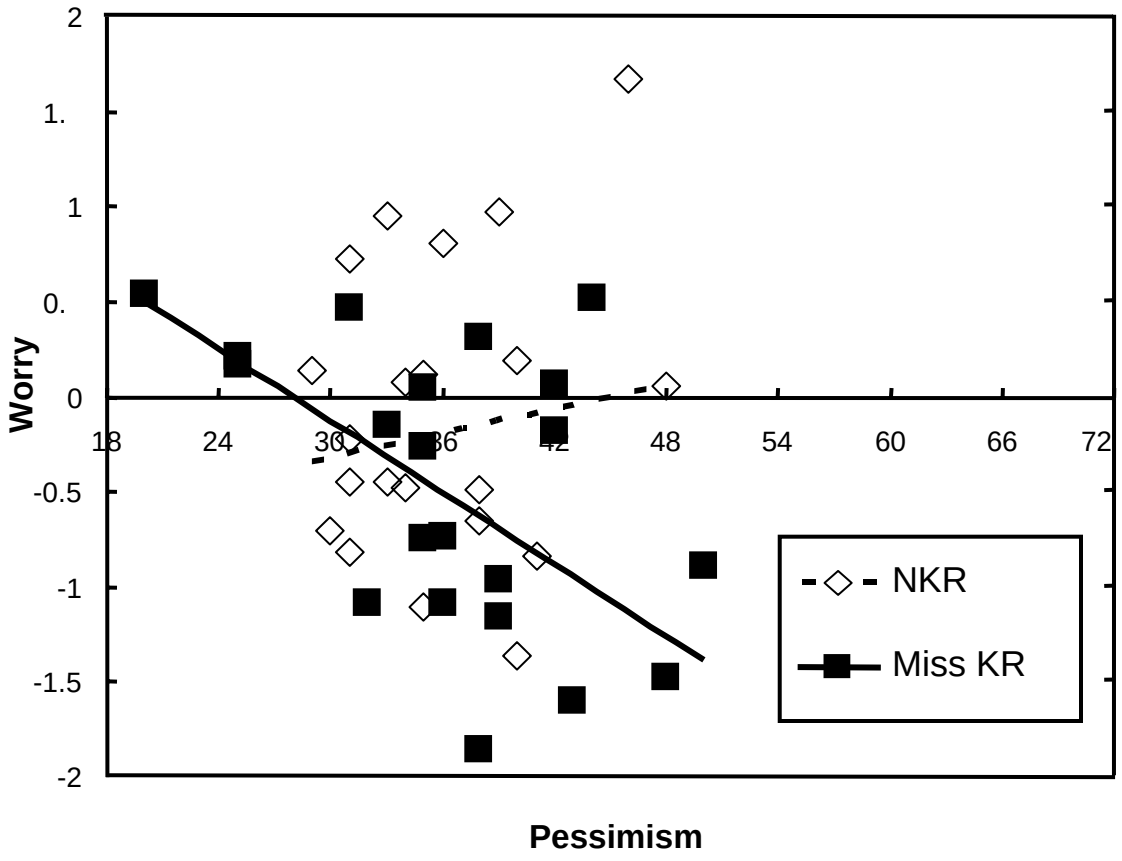

(a)
P
FI
FO
T
(b)
P
FI
FO
T
(c)
P
FI
FO
T
(d)
P
FI
FO
T
(e)
P
FI
FO
T
(f)
P
FI
FO
T
(g)
P
FI
FO
T

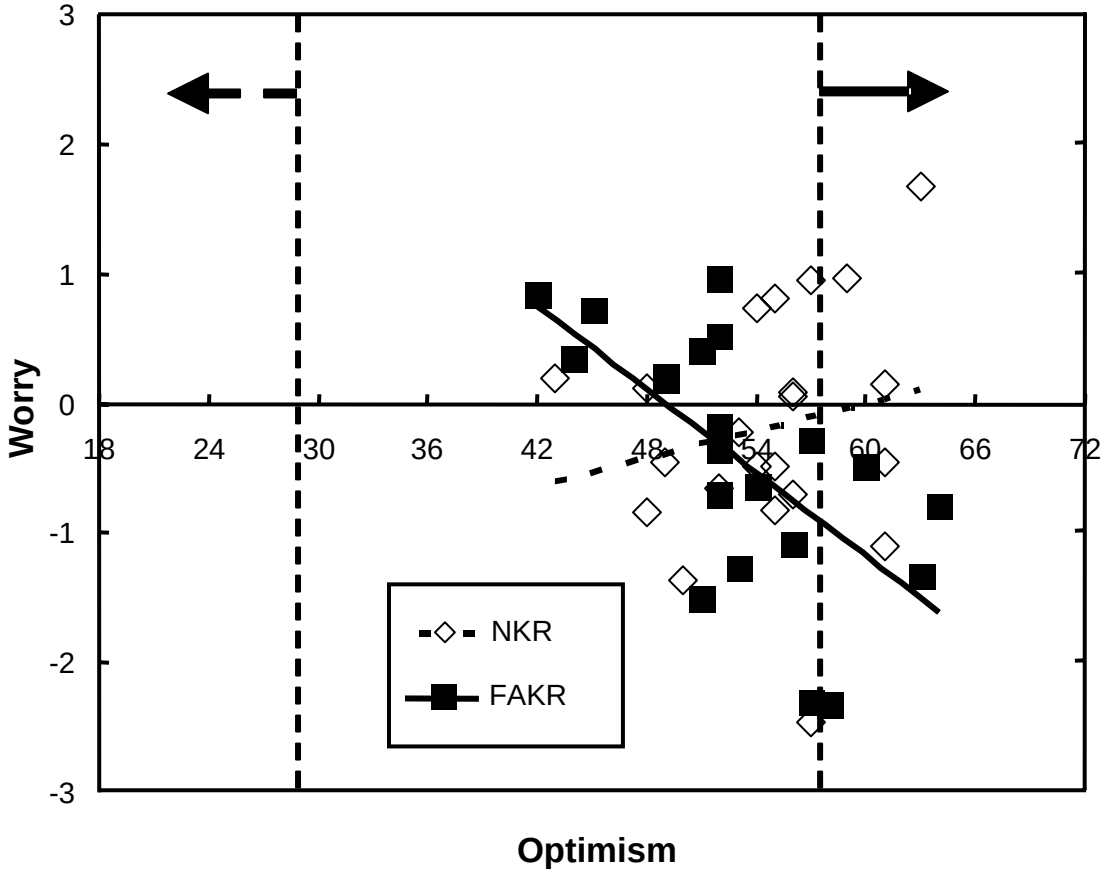

1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
A'
1
2
3
4
5
6
Period of Watch (5-min)
Composite KR
Hit KR
FA KR
Miss KR
NKR


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Bd"
1
2
3
4
5
6
Period of Watch (5-min)
Composite KR
Hit KR
FA KR
Miss KR
NKR


4
3
2
1
0
-1
-2
-3
-4
18
24
30
36
42
48
54
60
66
72
Distress
Pessimism
NKR
FAKR


3
2
1
0
-1
-2
-3
18
24
30
36
42
48
54
60
66
72
Task-Engagement
Pessimism
NKR
FAKR


2
1.
1
0.
0
-0.5
-1
-1.5
-2
Worry
18
24
30
36
42
48
54
60
66
72
Pessimism
NKR
Miss KR


3
2
1
0
-1
-2
-3
Worry
18
24
30
36
42
48
54
60
66
72
NKR
FAKR
Optimism

Remotely Operated Vehicles (ROVs) from the Top-Down and the Bottom-Up

T. Oron-Gilad[1], J.Y.C. Chen[2], and P.A. Hancock[1]

[1]University of Central Florida and [2] US Army Research Laboratory

Introduction

With the continuing evolution of technology, our capacity to sense environments at a distance has grown exponentially over the last two to three decades (Kurzweil, 2003). The natural outcome of this extension to sensory capabilities is the understandable wish to exert comparative action at a distance. But, imagine a world in which we could each effect action simultaneously at diverse locations! It would be like trying to have a phone conversation with every phone user in the world, all at the same time (after all, speech does effect action at a distance). Evidently, this would represent a chaotic cacophony in which, not only would you not be able to interact with everyone effectively, you would not be able to interact with anyone effectively. While extending our capacities for perception and action are evidently important, it is the selective and effective extension of such capacities we really desire.

ROV's are one obvious, contemporary technical instantiation of this evolutionary vector in extending human perception-action capabilities. As evident from our initial observations above, the central problem of ROV development is not the feasibility and creation of the hardware per se, as difficult an engineering challenge as this may be. Rather, the crucial issue is the assimilation of the relevant sensory inputs, the processing of information pertinent to specified user goals, and the translation of the user's

subsequent decisions into effective action. Thus, the fundamental barrier to success in this realm is not a technological one but a use-centered one (see Flach & Dominguez, 1995). In this chapter, we therefore adopt this use-centered perspective and present steps toward an integrated, functional architecture for ROV operations and eventually a much elaborated spectrum of advanced technologies. We do this by considering the issue from both a top-down and a bottom-up perspective.

The Top-Down Approach

When we take a top-down approach to understanding issues surrounding ROV implementation, we can employ the metaphor either literally or as a form of abstraction hierarchy (Rasmussen, 1986). Literally, the military's necessity for moment-to moment information mandates a suite of context-specific technological capabilities for sensor and effector systems. This suite includes but is not limited to systems in outer space (such as geo-synchronized orbiting platforms), high altitude atmospheric systems (such as Global Hawk) and other craft which operate less than hundreds of feet from earth down to almost ground level itself.

In the abstract sense, a top-down perspective requires us to be very explicit about the goals, the intentions, the requirements, the aspirations, and the limitations of the technology at hand. It is central to any abstraction hierarchy that we make explicit what we really require these burgeoning systems to accomplish. The military's necessity for moment-to moment information drives the development of technologies that will increase information gathering capacity. The development of manned flight is a good example. Flying is a continuous locomotion through a medium that does not permit unaided human

presence and thus conquering powered flight is viewed primarily as a technological achievement. Aircraft over the years have grown in their capacities as ever greater functionality was strapped to ever faster vehicles. The rationale for continuing pilot inclusion is always the "flexibility" they bring to the human-machine combination. The human aspiration for flight (see Hancock & Chignell, 1995; Magee, 1941) is often tainted by the pragmatic mandate which funds such emerging capacities. The intrinsic enjoyment of flight is now progressively ever more in conflict with the extrinsic reasons for military aircraft. It is into this milieu of dissonance that the advent of the military ROV is being thrust. In ROVs, the operator is included in the loop but the role of the operator and their motivation has is now very different. The intimate and physical relationship between pilot and aircraft is all but severed while the simple "love of flight" is, within these technologies, all but eliminated.

This brings us to the question of exactly what human capabilities are extended by ROV. In reconnaissance for example, the ROV acts predominately as an extension to the human eye. Whether that eye is a camera in near-earth orbit or a video on micro-light, the fundamental goal is the expansion of the range of the end user's perception. ROVs provide this 'eye in sky' capacity but the act of sensing is essentially divorced from the act of perception. This dissociation is highly problematic since the act of perception is itself an active one. Divorcing the process of sensory assimilation from perceptual interpretation and action engagement promises to fracture a fundamental human ability (Gibson, 1966; 1979; Powers, 1973). A frozen field of view provides only impoverished information. Normal vision is not compounded of snap shot-like units but is essentially

dynamic and is guided by previous actions taken. That is, perception is predicated on action just as much as action depends on perception. Little wonder then that "situation awareness" is seen as representing the crucial bottleneck in current systems (Endsley, 1995; Hancock & Diaz, 2001; Smith & Hancock, 1995). As ROV's increase in number, capacity, and form, we are rapidly overwhelming the ability of any one user, or even group of users, to extract useful and pertinent information. In a signal to noise sense, our noise level is becoming so high that no possible signal will be observable. In the dynamic conditions of combat, the problem is that signal can very rapidly turn to noise and vice-versa. Although the primary problem of today is one of information extraction, in the very near-term we will wish to effect action based on this derived information. If the problem is not solved as a perception-action issue there is good reason to believe it will not be solved at all.

Humans have always sought to extend their range of action commensurate with the range of perception. Indeed Hancock (2000) noted that "*the perception-action link explains how we explore, the perception-action gap explains why we explore*." Remotely operated combat vehicles (ROCVs) represent an obvious contemporary example of this technological magnification. Central limitations of user information assimilation, decision-making and response selection are thus the current barrier to achieving this goal. The top-down perspective emphasizes that the ROV is simply one of a suite of "vehicles" or technologies to achieve desired aims. In contrast, a bottom-up approach must take a much harder look at the limits and constraints present-day and near-term operational systems.

Humans have highly constrained sensory and effector bandwidths. We have evolved to tune these limited perception-action systems to the needs of survival and procreation. Now these peripheral capacities are magnified by technology but the comparable central capacities of perception, decision-making and response-selection remain largely unaltered. Technologies which seek to support such central capacities are fundamentally different from those that support peripheral capacities. In facing the harder challenge of cognition, they remain at present much less effective and so the former capacities remain the bottleneck of concern. It is anticipated that the future fighting personnel will have to function under very compressed planning and operating cycles and at very high tempos (Scales, 2001). Under these stresses, the importance of decision-support expressed in the human-system interface will grow ever greater, making it even more crucial to frame information exchange so as to best facilitate human decision and action response (and see Hancock, Flach, Caird, & Vicente, 1995; Mouloua, Gilson, & Hancock, 2003).

Recently we have argued that the ROV control ratio (i.e., how many ROVs can a single operator control) is not the right question (Hancock et al., 2005). We now wish to expand upon this concern. We argue that under extreme stress, any single operator can and should gain effective situational awareness of only one relevant operational environment at a time. Under stressful conditions, individuals report a phenomenological 'narrowing' of the range of cues they extract from the surrounding environment both spatially (Easterbrook, 1959) and temporally (Hancock & Weaver, 2005). The vast amounts of data that remotely operated vehicles provide render these operators immediately vulnerable to the effects of information overload. Thus, when we see

technology as the answer, the end point of that general strategy is an evident situation in which technology becomes the central problem. The tasks associated with use of ROV's include inherent spatial and temporal uncertainty in regard to targets, which result in extended periods of underload (i.e., a vigilance imperative), as well as the sudden overload associated with the onset of conflict and as well as the transitions between these two states (Huey & Wickens, 1993). Furthermore, these effects vary with the context of operation. In addition, increased use of automation can cause significant and rapid fluctuations in operator workload (e.g., Scallen, Duley, & Hancock, 1996). The key to successful implementation of this technology for use in stressful environments is to ensure that the information presented to operators is structured in manner amenable to direct interpretation with a minimum of cognitive processing demands (Hancock & Szalma, 2003). If top-down ideas differentiate the feasible from the possible, bottom-up approaches seek to describe the actual from the highly probably. We now turn to this bottom-up perspective.

The Bottom-Up Approach

The space-time taxonomy, initially introduced by Ellis for network categorization (Ellis, Gibbs, & Rein, 1991) captures the basic categorization of the physical and temporal relations between the machines and the humans. This categorization has already demonstrated the capacity to impact on automation implementation (see Yanco & Drury, 2002). This space-time taxonomy divides user-machine interaction into four categories based on whether the humans or the machines are using their computing systems at the same time or at different times and while in the same place or in different places. What

this taxonomy lacks is the literal reference to the time scale of relevance and the distance of physical proximity.

We therefore suggest a broader space and time taxonomy where space is defined in two dimensions: physical size (nano to macro) and physical proximity, and time is also defined in two dimensions: time scale (momentary and extended) and synchronization. This enhancement of the space-time taxonomy is critical for identifying potential situation awareness limitations and important for the implementation of future ROV's capacities such as adaptive automation. Action frequently provides real-time response feedback, letting the human know the outcome of their efforts in order to match outcome to intention. User performance is vastly disrupted if this feedback loop is distorted either spatially or temporally (Smith, 1993; Smith & Smith, 1993). Each spatial and temporal dimension affects the operator's ability to perceive information and can have an immediate influence on the effectivity of the feedback loop, particularly when action at distance is involved.

Spatial context is critical for situation awareness and feedback as it defines the ability of the human operator to understand environmental elements related to the ROV and enhances task engagement. Teleoperation is often prone to poor spatial awareness of the remote environment (see for example, Darken & Peterson, 2002; Tittle, Roesler, & Woods, 2002). Also, the use of cameras to capture remote environments frequently create a "keyhole" effect or "soda straw" view (Woods, Tittle, Feil, & Roesler, 2004; Murphy, 2004). Furthermore, if the spatial scale is very small (e.g., nano-scale) or very large (e.g., thousands of kilometers) it becomes difficult for the operator to relate empathetically to

the perceived entities in terms of personal experience (and see Hancock & Chignell, 1995). Physical proximity determines how close the human and the machine are; high proximity means that the human and the machine share the same local environment (i.e., the same 'workplace'). It implies that the operator is likely to have an immediate empathy with local conditions. In high proximity situations the physical laws that apply to the system are similar to those that apply for the operator. Progressively lower values of proximity indicate that there is a physical distance (or separation) between the operator and the machine. Hence, a nano-scale may be similar in proximity distance-value to a macro-scale situation because each are equivalently different from common human experience. This proximity is not a simple spatial referent per se. Lower proximity increases the need to provide more situational information of that remote environment. For instance, Thomas and Wickens (2000) found "cognitive tunneling" when viewing a remote environment using an immersive displays (such as the ones typically used for ground robots) compared to displays with exocentric frame of reference (similar to views from a UAV), which had a greater FOV. Furthermore, humans tend to underestimate distances more in photographic quality imagery than in the real world (Lampton, Singer, McDonald, & Bliss, 1995; Witmer & Kline, 1998; Thompson et al., 2004).

Traditionally, time has been associated with reaction time, latency of cognitive responses (e.g., Card, Moran, & Newell, 1980), and motor movement (e.g., Fitts, 1954). Recent studies (e.g., Lane et al., 2002; MacKenzie & Ware, 1993; Malcolm & Lim, 2003) reinforce the problematic operator-machine coupling when time delays occur. Sheridan (2002) recommended that supervisory control and predictor displays be used to

ameliorate the negative impact of time delays on teleoperation. Furthermore, for the implementation of adaptive automation, the understanding of human perception of time, particularly under stress, becomes a critical consideration (see, Hancock, & Scallen, 1996).

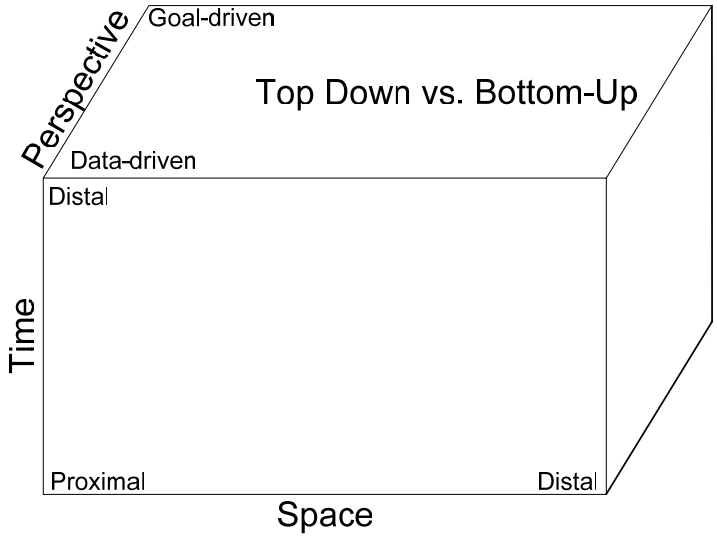


Figure 1. A descriptive three dimensional model which incorporates the operational perspective with the space-time taxonomy.

What remains yet to be determined is how to frame information exchange to best facilitate human decision and action response within this remote perception-action system. Figure 1 provides a descriptive three dimensional model which incorporates the operational perspective with the space-time taxonomy. The operational perspective can be either goal-driven or data-driven. There are commonalities between a goal-driven and a top-down abstraction and vice versa; there are commonalities between the data-driven approach and a bottom-up view point. The data driven perspective is a more passive situation where the operator merely responds to events in the environment if they occur

(see also Fong et al., 2004). From a data-driven perspective our expectation is that, the 'level' at which the individual (inter)actor engages with the system, represents the crucial arbiter on performance. Those in greatest proximity to the theater of operations are most liable to respond under conditions of stress and fatigue. They are therefore more likely to adopt a data-driven (bottom-up) approach, as immediate changes in threat or identification of unknown objects may have a direct effect on their safety and the safety of their unit of action. In such circumstances the immediate goal changes from the higher level tactical operational goals to a more self-centered safety goal. More distal operators are liable to have access to facilities which will mitigate such effects (although in times of great demand such as combat they will be omni-present in the theater of operations). The goal-driven strategic recommendations will depend upon the hierarchy of operational contexts. Close to the scene of action, we have previously distilled that simple, graphical representations with unambiguous cues for response need to be employed. As the individual becomes more distal to the locus of action, the imperative to provide these simplifications decreases. Table 1 addresses issues of concern in regard to each dimension of the model. This table is only an exemplar and additional aspects such as communication and coordination will probably be added based on the context of operation.

Table 1

Issues of concern in regard to each dimension of the model

Space: high - low proximity
• Limited View ("keyhole" effect) • Restricted Field of View: affects target detection, distance estimation, and identification of self location • Degraded Depth Perception: affects estimates of distance and size • Camera Viewpoint (Context): contextual change when switching among robots • Degraded Video Image**:** affects distance, object identification, and size estimation • Reduced Bandwidth: affects spatial orientation, object identification, and speed and motion perception
Time: delay and magnitude
• Compensatory tracking performance degrades with a latency of about 300 ms. • When latency is over 1 second, operator switches to a "move and wait" control strategy • Placement task performance degrades when standard deviation of latency is above 82 ms • Over actuation is common when system delay is unpredictable
Perspective: bottom-up (data driven) - top-down (goal driven)
Bottom-up: passive perception (interpretation of sensor data) • Effectiveness and efficiency of sensor data retrieval • Cognitive tunneling *Top-down: active perception (seeking sensor data)* • Effective ways to present information and provide decision support aids • Span of control • Intent or directedness toward future state or goal

Adaptive automation helps to manage momentary changes and thereby efficiently schedule tasks and regulate demands. However, understanding the reciprocal relationship between adaptive automation and operator performance is limited by the sparse level of systematic research and a still under-developed theoretical framework (although see Hancock & Chignell, 1987; Parasuraman, Sheridan, & Wickens, 2000). Though we recommend and advocate adaptive automation at all levels of the system, interface appearance will change throughout the system, according to the operation, operator, and ROV characteristics. Although inter-operability remains an important issue, and the minimization of training time a real concern, our expectation at present is that multiple

levels of operational sophistication will mandate hierarchic and selectable interface modes.

Discussion and Conclusions

There is a very real danger when we equate progress directly with increasing technical capacity (Perrow, 1984). In actuality, progress is less and less accurately indexed by technical capacities as it is by the ability to achieve the human goals for which it was designed. ROV's are a cusp technology in this transition of understanding. Almost inevitably, we focus on flight envelopes, operational capacities, physical limitations and the like. The overwhelming majority of resources are thrown at these problems in the desperate search for solutions. However, if we start from what it is the user needs then we need not seek optimal engineering solutions and indeed may circumvent some problematic technical issues altogether. For the fundamental problem here is information not technology. Our brief take on system architectural issues proceeds from this premise but like all in the human sciences we are laboring against the perception of the problem as opposed to the true requirements. On the programmatic level, the issues that need immediate attention include information sharing, fostering situation awareness, multiple control problems, display design and hardware barriers to achieve effective early prototypes. Several of those issues especially that of human-in-the-loop and semi-automated control are the subject of progressing research (Sheridan, 2002). The issue of control is a vital one since it is evident that we still design as though the machine system is to react conceptually and cognitively as 'one of us'. For example, we still seek explicit extensions to human sensory capabilities while the integration of

other sensory information still lags behind. Like most such situations, we will have to rely on the rebound from failures, some of which have already occurred. In the end technologies which fail to achieve the desired aim will fade away. How and when this evolution will occur depends directly on the foresight of ROV designers.

Acknowledgments

This work was supported in part by the Army Research Laboratory through the Micro-Analysis and Design CTA Grant DAAD19-01C0065, Michael Barnes, *Technical Monitor*, P.A. Hancock, *Principal Investigator*. This work was also facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, Dr. Elmar Schmeisser, *Technical Monitor*, P.A. Hancock, *Principal Investigator*. The views expressed in this work are those of the authors and do not necessarily reflect official Army or Department of Defense policy.

References

Bryant, D.J., Lichacz, F.M.J., Hollands, J.G., & Baranski, J.V. (2004). Modeling situation awareness in an organizational context: military command and control. In: S. Banbury and S. Tremblar (Eds.) *A cognitive approach to situation awareness: theory and application.* Ashgate: Burlington, VT.

Card, S.K., Moran, T.P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. LEA: Hillsdale NJ.

Darken, R. P., & Peterson, B. (2002). Spatial orientation, wayfinding, and representation. In Stanney, K. (Ed.), *Handbook of Virtual Environment Technology* (pp. 493-518). Mahwah, NJ: Erlbaum.

Easterbrook, J.A. (1959) The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, **66**, 183–201.

Ellis, C.A., Gibbs, S.J., & Rein, G.L. (1991). Groupware: some issues and experiences, *Communications of the ACM*, **34**(1), 39-58.

Endsley, M.R. (1995). Towards a theory of situation awareness. *Human Factors*, **37**(1), 32-64.

Flach, J.M. & Dominguez, C.O. (1995). Use-centered design: Integrating the user, instrument, and goal. *Ergonomics in Design*, **3**(3), 19-24.

Fitts, P.M., (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, **47**(6), 381-391.

Fong, T., Kaber, D., Lewis, M., Scholtz, J., Shultz, A., & Steinfeld, A. (2004). Common metrics for human-robot interaction. In *Proceedings of IEEE International Conference on Intelligent Robots and Systems* (Sendai, Japan). Retrieved February 23, 2004, from http://vrai-group.epfl.ch/papers/IROS04-TF.pdf

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Hougton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception***.** Boston, MA: Houghton Mifflin.

Hancock, P.A. (2000). Presidental address. In: P.A., Hancock (Ed.), (2005). *The human face of technology: Selected Presidential Addresses of the Human Factors and Ergonomics Society*. Human Factors and Ergonomics Society: Santa Monica, CA.

Hancock, P.A., & Chignell, M.H. (1995). On human factors. In: Flach, J., Hancock, P.A., Caird, J.K., and Vicente, K. (Eds.). *Global Perspectives on the Ecology of Human-Machine Systems.* Lawrence Erlbaum: Mahwah, NJ.

Hancock, P.A., & Chignell, M.H. (1987). Adaptive control in human-machine systems. In P.A. Hancock (Ed.), *Human factors psychology* (pp. 305-345). North Holland: Elsevier Science Publishers.

Hancock, P.A., & Diaz, D. (2001). Ergonomics as a science of purpose. *Theoretical Issues in Ergonomic Science*, **3** (2), 115-123.

Hancock, P.A., Flach, J., Caird, J.K., & Vicente, K. (Eds.) (1995). *Local applications in the ecology of human-machine systems*. Lawrence Erlbaum: Mahwah, NJ.

Hancock, P.A., Mouloua, M., Gilson, R.D., Szalma, J.L., & Oron-Gilad, T. (2005). Is the UAV control ratio the right question? Paper submitted.

Hancock, P.A., & Scallen, S.F. (1996). The future of function allocation. *Ergonomics in Design*, **4** (4), 24-29.

Hancock, P.A., & Szalma, J.L. (2003). Operator stress and display design. *Ergonomics in Design*, **11** (2), 13-18.

Hancock, P.A., & Weaver, J.L. (2005). Temporal distortions under extreme stress. *Theoretical Issues in Ergonomic Science*, **6** (2), 193-211.

Huey, M.B., & Wickens, C.D. (Eds.). (1993*). Workload transition: Implications for individual and team performance*. Washington, D.C.: National Academy Press.

Kurzweil, R. (2003). *The Law of Accelerating Returns.* Retrieved July 7, 2005 from www.KurzweilAI.net

Lampton, D.R., Singer, M.J., McDonald, D.P., & Bliss, J.P. (1995). Distance estimation in virtual environments. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1268-1272). Santa Monica, CA: Human Factors and Ergonomics Society.

Lane, J.C., Carignan, C.R., Sullivan, B.R., Akin, D.L., Hunt, T., & Cohen, R. (2002) Effects of time delay on telerobotic control of neutral buoyancy vehicles. In *Proceedings of IEEE International Conference on Robotics and Automation.* Retrieved February 13, 2004, from www.ssl.umd.edu/publications/SSL02-003.pdf

MacKenzie, S., & Ware, C. (1993). Lag as a determinant of human performance in interactive systems. In *Proceedings of ACM Conference on Human Factors in Computing Systems- INTERCHI '93* (pp. 488-493). New York: NY: ACM SIGCHI. Retrieved May 27, 2004, from http://www.yorku.ca/mack/p488-mackenzie.pdf

Magee, J.G. (1941). High Flight. http://www.wpafb.af.mil/museum/history/prewwii/jgm.htm

Malcolm, A.A., & Lim, J.S.G. (2003). Teleoperation control for a heavy-duty tracked vehicle. Retrieved June 21, 2004, from http://www.simtech.a-star.edu.sg/Research/TechnicalReports/TR0369.pdf

Mouloua, M., Gilson, R., & Hancock, P.A. (2003). Designing controls for future unmanned aerial vehicles. *Ergonomics in Design*, **11** (4), 6-11.

Murphy, R.R. (2004). Human-robot interaction in rescue robotics. *IEEE Systems, Man and Cybernetics Part C: Applications and Reviews, special issue on Human-Robot Interaction*, *34*(2). Retrieved Apr. 16, 2004, from http://crasar.csee.usf.edu/research/Publications/CRASAR-TR2004-2.pdf

Parasuraman, R., Sheridan B.T., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics: part A* 30(3), 286-297.

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies.* NY: Basic Books.

Powers, W.T. (1973). *Behavior: The control of perception*. Chicago: Aldine.

Rasmussen, J. (1986). *Information Processing and Human-Machine interaction: An Approach to Cognitive Engineering*, New York: North-Holland.

Scales, R.H. (2001). *Future warfare anthology*. US Army War College, ISBN 1-58487-026-5.

Scallen, S.F., Duley, J.A., & Hancock, P.A. (1996). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics*, **26**, 397-403.

Sheridan, T.B. (2002). *Humans and Automation: System Design and Research Issues.* Hoboken, NJ: John Wiley & Sons.

Smith, K., & Hancock, P.A. (1995). Situation awareness is adaptive, externally-directed consciousness. *Human Factors*, **37**, 137-148.

Smith, T.J. (1993). Automation of work in dangerous environments. In: M.J. Smith and G. Salvendy (Eds.). *Human-Computer Interaction: Applications and Case Studies, Proceedings of the Fifth International Conference on Human-Computer Interaction*, Volume 1, pp. (273-277), Elsevier.

Smith, K.U., & Smith, T.J. (1993). Human-computer interaction and the automation of work. In: M.J. Smith and G. Salvendy (Eds.). *Human-Computer Interaction: Applications and Case Studies, Proceedings of the Fifth International Conference on Human-Computer Interaction*, Volume 2, pp. (837-842), Elsevier

Thomas, L.C. , & Wickens, C.D. (2000). *Effects of Display Frames of Reference on Spatial Judgments and Change Detection.* (Tech. Report: ARL-00-14/FED-LAB-00-4). Urbana-Champaign, IL: University of Illinois, Institute of Aviation, Aviation Research Laboratory. Retrieved April 4, 2005, from http://www.humanfactors.uiuc.edu/Reports&PapersPDFs/TechReport/00-14.pdf

Thompson, W.B., Willemsen, P., Gooch, A.A., Creem-Regehr, S.H., Loomis, J.M., & Beall, A.C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence, 13,* 560-571.

Tittle, J.S., Roesler, A., Woods, D.D. (2002). The remote perception problem. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 260-264). Santa Monica, CA:

Human Factors and Ergonomics Society.

Witmer, B.G., & Kline, P.B. (1998). Judging perceived and traversed distance in virtual environments. *Presence, 7*(2), 144-167.

Woods, D. D., Tittle, J., Feil, M., & Roesler, A. (2004). Envisioning human-robot coordination in future operations. *IEEE Transactions on Systems, Man & cybernetics: Part C: Special Issue on Human-Robot Interaction.* Retrieved April 21, 2004 from http://csel.eng.ohiostate.edu/woods/papers/woods_hri_IEEE.pdf

Yanco, H.A. & Drury, J.L. (2002). A taxonomy for human-robot interaction, *AAAI Fall Symposium on Human-Robot Interaction, AAAI Technical Report FS-02-03*, pp. 111-119, November.

The Magazine of Human Factors Applications

VOL 13, NO 1

ERGONOMICS IN DESIGN

WINTER 2005

Hedonomics: The Power of Positive and Pleasurable Ergonomics

ALSO IN THIS ISSUE:

Predicting Product Misuse

Road Safety Signs



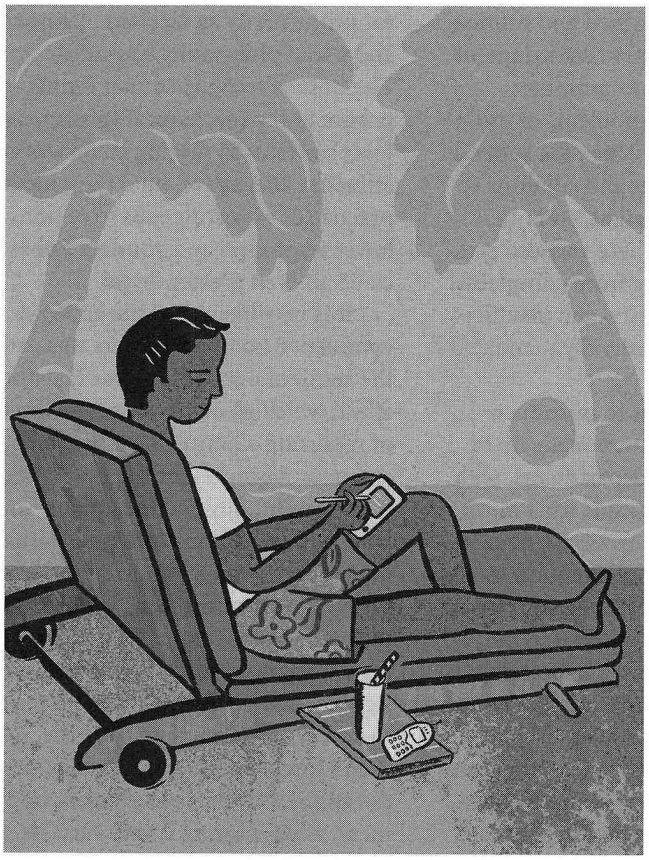
Non Profit Org.
U.S. Postage
PAID
Permit No. 4
Hanover, PA

Human Factors and Ergonomics Society

Hedonomics:
The Power of Positive and Pleasurable Ergonomics

BY PETER A. HANCOCK, AARON A. PEPE, & LAUREN L. MURPHY

Once individual users' safety and functionality needs are satisfied, pleasure should be considered in the design of products, systems, tools, and environments.



In this article, we introduce a new term into the human factors/ergonomics lexicon: *hedonomics.* We define hedonomics as that branch of science and design devoted to the promotion of pleasurable human-technology interaction. In advocating for hedonomics, we seek to augment and expand ergonomists' arsenal for improving the design of all human-machine technology.

Our intention here is to (a) provide a philosophical framework for hedonomics; (b) contribute a model of a design priority hierarchy, in which we explain in detail the concept of individuation as the ultimate human approach to technology; and (c) offer potential design guidelines for and areas of hedonomic research that we believe should be pursued as an agenda for progress in the near future.

GENESIS AND MEANING OF *HEDONOMICS*

Like the term *ergonomics*, hedonomics is derived from two Greek roots: *hedon(e)*, meaning joy or pleasure, and *nomos*, meaning law-like or collective. In ergonomics, our traditional moral ethos is founded on the praiseworthy efforts to prevent people's pain and suffering, predominantly in the workplace. These concerns have grown over recent decades into a formal branch of science now devoted to a wider set of improvements in all physical and cognitive environments, thus serving to enhance the overall quality of life.

In contrast to the prevention of pain, but much in the same realm of human-technology interaction, hedonomics is primarily concerned with the promotion of pleasure. Hedonomics and human factors/ergonomics (HF/E) are consequently two sides of the same coin. However, whereas HF/E is often handicapped by trying to show the value of preventing events that eventually do not happen, hedonomics, being a positive enterprise, rejoices in the advantage of showing the value of events that do.

Fundamentally, ergonomics and hedonomics are synergistic perspectives directed toward the same goal of optimized human-technology interaction, which is central to the goal of all future design. Perhaps our first responsibility is to clarify an important issue. Although one of us (Hancock) coined the term *hedonomics*, the pioneering work in this area is primarily that of others (for example, Blythe, Overbeeke,

Copyright 2005 by Human Factors and Ergonomics Society, Inc. All rights reserved.

Monk, & Wright, 2003; Jordan, 2000). With these and other like-minded individuals, we see great value in this form of positive ergonomics, and this article articulates our current conceptual perspective.

Given that hedonomics, like HF/E, is a hybrid, practical, and interdisciplinary area, the base knowledge is drawn from multiple fields in the sciences and humanities. Thus, hedonomics in its early form draws heavily on extant discipline-based constructs varying from examples such as emotion, or affect (Snyder & Lopez, 2002). However, the expectation is that it will develop as an individual and distinctive area of study. Many readers may already be familiar with the general concept of designing with pleasure in mind.

From Tradition to Innovation

Traditionally, design has conceived and created systems that largely appease the demands of the neocortex. This is reasonable given that the pragmatic needs of functionality mandate a foundation based on purely rational processes. However, in so doing, we have largely neglected the needs of the limbic system. Much appeal has been made to design for efficient, speeded, error-proof (or at least error-resistant) systems with the fundamental underlying ethos that the human is a part of the final system in the same way that other inanimate components are also parts of that same system. Even from this uncompromising original conception, based in an overwhelming concern for process efficiency, it has eventually emerged that a so-called human-centered strategy provides a superior outcome as an overall design ethic, especially as technical systems increase in complexity. Human-centered design places the concerns for the most fragile, the most unpredictable, but the most crucial component (the human) at the very center of system conception and thus temporally at the earliest initiation phases of design.

But in what sense has human-centered design been human? Concern is not expressed for the desires of *each individual person* but for the ways in which understanding the general sensory, cognitive, and response capacities of *any human* can be used to best advantage. In this sense, one could as easily generate monkey-centered design or, as evident from the work of Pepperberg (1999), parrot-centered design! There is a wonderful creative article on how to design for centaurs (Wise & Fey, 1981). Centaurs are mythical creatures, it is true, but centaur-centered design is a feasible and achievable proposition within the overarching conception that makes the living system the central focus.

FEATURE AT A GLANCE: Hedonomics takes ergonomics one step forward from its initial prevention of pain to its contemporary promotion of pleasure in human-technology interaction. We provide a philosophical framework, contribute a model of a design priority hierarchy, and offer potential areas of future research for hedonomics, which we define as the promotion of pleasurable human-machine interaction. Individuation and experience goals should be incorporated into fundamental design priorities in order to optimize user experience and improve the quality of life for all affected by advancing technology.

KEYWORDS: Well-being, affect, performance capability, hedonic design, happiness, usability

However, there are many unique dimensions to each individual human being (as perhaps there are to each monkey, parrot, and centaur). These characteristics extend beyond the common, shared sensory and cognitive capacities or momentary response limitations. Such dimensions include self-motivation, affect, pleasure, or, more generally, the idiographic apperception of the world. Recognition of the importance of these personal dimensions encourages us to explore their impact and evaluate each individual with respect to his or her own unique motivations and aspirations.

Starting from the premise that emphasizes commonalities across all human beings, we seek to extend such concern to an individual-based design.

The continuing revolution of human-centered design of the latter part of the twentieth century is evident in the gradual recognition of this primacy of human capacities as the centerpiece of design (see Billings, 1991). Although it is true that this principle has yet to thoroughly penetrate the full spectrum of design, its logical evolutions have already begun to emerge. For example, Flach and Dominguez (1995) emphasized the importance of goal achievement in the principle of "use-centered design." Here the key element is *use*, which is a property that emerges from the combinatorial actions of humans and technology, rather than being a unique property of either.

This article takes human-centered design in a different but equally logical direction. Starting from the premise that emphasizes commonalities across all human beings, we seek to extend such concern to an individual-based design or state of *individuation* in which the focus is on the unique characteristics of each and every single individual user (e.g., Peter vs. Aaron vs. Lauren). We believe technology has now progressed to a point that it is capable of dealing with users on this individual, personal level, and this progress fits hand-in-glove with the growing concerns for affect as embraced by hedonomics.

Individuation is not just the pursuit of self-centeredness. This is so because optimization of an interface for a specific user alone, which allows him or her to interact intuitively and personally, does not affect only the user; it also affects the user's environment and all aspects of that environment, including other people (see Hancock, 2003). This widespread effect could be either pleasant or detrimental in nature, given that satisfying one individual's hedonomic need has no necessary link to the hedonic state of any other individual. For example, Moray (1993) commented on the benefit of smart houses for "making life easier and safer for inhabitants

and reducing energy consumption." The ability for a smart house to detect and adapt to individual characteristics of its inhabitants would not only allow for individual comfort and optimal operation but also could conserve energy for the greater good – in this case, a collective positive benefit (see Mynatt, Melenhorst, Fisk, & Rogers, 2004).

Additionally, human factors/ergonomics specialists have a responsibility to consider potential negative effects of the outflow of poor design and the failure that is produced when we do not take into consideration the strengths and limitations of each human (Vicente, 2004). A bitter but pertinent example derives from the human factors failures that led to disaster at Chernobyl, which affected not only the lives of the exposed controllers but also the individuals in the surrounding communities and eventually people of many other nations.

From Preventing Pain to Promoting Pleasure

From its formal inception in the middle of the nineteenth century (Jastrzebowski, 1857) to its flowering in the post-World War II era (Marek & Pokorski, 2004), ergonomics has been founded on the fundamental and morally laudable concept of the prevention of pain. Whether this prevention addresses acute effects in terms of bodily injury or the more chronic expressions involved with repetitive strain trauma to either musculature or cerebrum, ergonomics seeks to dissipate the entropy evident in maladaptive operational conditions (also see Karwowski, 2000). This goal is achieved largely through design and redesign of aspects of the surrounding environment.

Frustration, pain, stress, fatigue, overload, injury, and death are all symptoms of systems that desperately require the attention of the professional ergonomist. Preventing bad things from happening through design, redesign, or other remedial processes is an activity we share with a larger community of professionals such as those in safety and occupational health. However, all preventative measures are acts of negation. Logically, attempts to value such acts come under the general compulsion of proving a negative, which remains both a philosophical and a practical conundrum. We refer to this historical and traditional foundation as the *ergonomics of negation.*

Human factors/ergonomics is not the only scientific enterprise that has focused on prevention or subsequent damage limitation; it is only one of many sciences whose primary aim has been to seek the reduction of adverse events. Psychology, one of the base disciplines of HF/E, has been especially focused in the post-World War II era on addressing the problems of disturbed and mentally infirmed individuals and those who are suffering psychological illness. Even many so-called pure research efforts in the supporting neurosciences are founded on the desire to push for knowledge to alleviate neurological problems; thus, much of the contemporary focus on drug therapies. However, rather than avoiding or trying to cure the negative, there is a completely different perspective that focuses on the promotion of the positive.

Accentuate the Positive

Like a number of other human-based disciplines, human factors/ergonomics has now begun this transition from a basic concern about the issues of pain and injury to a greater focus on the pleasurable aspects of interaction (see Helander, Khalid, & Tham, 2001). This potential paradigm shift – or, more correctly, paradigmatic expansion – is evident in many growing spectra of human science (e.g., Suedfeld, 2001). Indeed, the aims of so-called positive psychology coalesce completely with those of hedonomics because both innovations seek to improve different but related facets of the overall quality of life.

> To fulfill the needs of the user, we need to incorporate an explicit recognition of motivation, quality of life, enjoyment, and pleasure into design recommendations.

Peterson and Seligman (2004) approached the cultivation of individual human strengths and character by developing a taxonomy of character virtues that includes wisdom, knowledge, courage, humanity, justice, and temperance. This work is part of a general effort by leaders such as Seligman and Csikszentmihalyi (2000) who have appealed for this reorientation of focus. They advocate that attention should now be directed toward an understanding of the benefits rendered by psychological insights into well-adjusted, so-called normal individuals for whom we can promote positive life experiences. These include explanations of a sense of expansiveness or awe (Woodruff, 2001) and the notion of optimal experience or "flow" (Csikszentmihalyi, 1990).

This positive vision transfers especially well to the practice of hedonomics in design. In the past decade, we have seen progress in human-computer interaction (HCI) and usability that has begun to be similarly oriented in this positive fashion (see Teague & Whitney, 2002). This development has proceeded along a path from an initial focus on errors and failures to today's considerations for user preference. Contemporary efforts (Han & Hong, 2003; Helander & Tham, 2003; Khalid, 2004; Krippendorff, 2004; Liu, 2003) are directed toward the search for the promotion of positive, pleasurable experience to augment and expand on the laudable efforts to prevent negative outcomes.

Unfortunately, most usability studies that integrate user experience have focused largely on aspects of user satisfaction or preference as a measure of their immediate success. The reification of any form of subjective response often leaves a field largely bereft of theoretical foundation and sometimes even scientific grounding. Given that satisfaction can be, and most often is, a predominantly momentary or transient state that is unique to the individual, hedonomics is challenged with the same scientific developmental issues as usability, positive psychology, or any enterprise in which private experience is a central element (see Lazarus, 2003).

Hedonomic Hierarchy of Needs

In order to establish hedonomics as a field of study, we provide an initial theoretical framework through which to promote the opportunity for pleasurable interaction based on a design priority hierarchy. In human-machine interaction, the system must first be designed to be safe, functional, and usable before it is designed to be pleasurable. Maslow's (1970) model of the optimization of human satisfaction through a hierarchy of needs specifies that higher-level needs can be fulfilled only after lower-level needs are satisfied. We take this principle as a design imperative and for our ergonomic and hedonomic hierarchy of needs, which is shown in the illustration below.

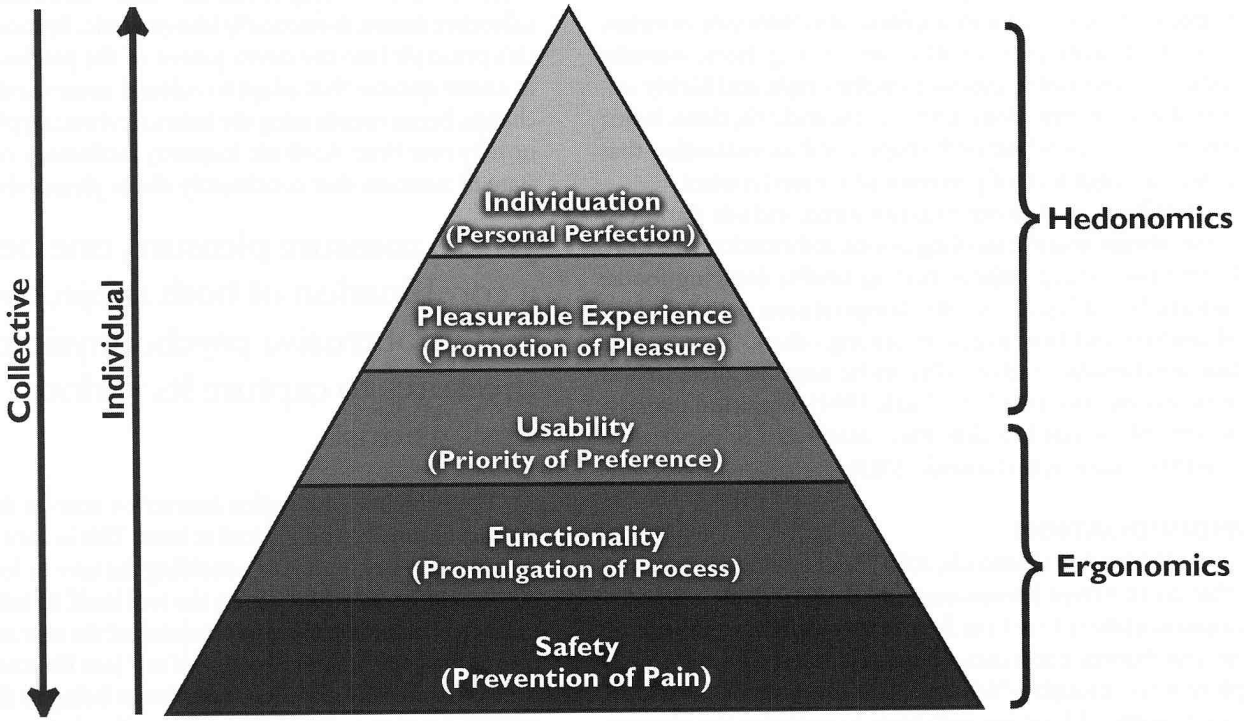
The foundation of our hierarchy is safety. Those who are in peril cannot be expected to reach for or be concerned with either pleasure or individuation. Thus, the necessity to ensure the well-being of the user is a mandatory requirement that must be achieved before subsequent design goals are considered. Once the user is ensured of safe operating conditions, the next level requires a functional system that enables the user to accomplish his or her desired goal. Safety and functionality are twin basic needs on which high-level aspirations are based.

Once functionality is achieved, usability facilitates performance by consistently enabling the user to accomplish the task. This experience conditions the user to develop a sense of trust; he or she can perceive the system as consistently stable, predictable, and safe. In addition, usability improves the system by making it more memorable, learnable, efficient, and easy to use, which are base elements for pleasurable interaction. At these three levels of safety, functionality, and usability, design is human-centered or nomothetic in nature.

When the system has achieved functionality and usability, it then can be designed to fulfill users' more fragile psychological and sociological needs, such as their need to belong, to achieve, to be competent and independent, thus making interaction a fully pleasurable experience. To fulfill the needs of the user, we need to incorporate an explicit recognition of motivation, quality of life, enjoyment, and pleasure into design recommendations. Designing the system to aid the users in the need to live up to their fullest and unique potential is the paramount goal of hedonomics. This final goal can be accomplished through individuation, which incorporates customization principles to permit each individual to adapt the system to his or her individual traits, personal idiosyncrasies, changing goals, and mood states over time.

From Functional Usability to Hedonomic Fulfillment

Usability has typically been defined as seeking three goals: effectiveness, efficiency, and user satisfaction (Preece, Rogers, & Sharp, 2002). Of these three, the first two have received the most attention. It seems that satisfaction has taken the place of the user in general experience in interacting with the system; it is rarely measured as anything more



A hierarchy of ergonomics and hedonomic needs derived from Maslow's conception. The fact that these design imperatives match the social edict of "life, liberty, and the pursuit of happiness" has not escaped our attention (see Hancock, 1999).

than user preference. Moreover, when satisfaction issues conflict with the measures of effectiveness and efficiency, they are most often ignored. Indeed, many of the most popular surveys for usability (for example, QUIS) have completely left out any reference to satisfaction, much less pleasure. The reason that user satisfaction has been overlooked is partly because of insufficient measurement specifications and implementation guidelines. Hedonomics aims to develop these and other metrics of user experience.

In order to fulfill the expectations of the user, we need to incorporate experience goals as primary goals at the level of fundamental system design and development. Traditionally, experience goals have been left up to graphic designers, marketers, and advertisers (Teague & Whitney, 2002). Such research relies on focus groups and samples of opinion as to what are preferred interactive modes. Intrinsic to such an evaluative procedure is a concern for enjoyable and pleasurable interaction, but it is frequently expressed as summated opinion of the selected sample group. As we will show, the emphasis on the individual at the idiographic level can provide pleasure-promoting systems but ones that are unanchored in a world of subjectivity. In order to incorporate experience goals at the fundamental level, a scientifically grounded framework needs to be developed. One start toward developing this framework is through the conception of *hedonomic affordances.*

Hedonomic affordances invite an emotional reaction from the user by way of the objects' physical attributes that result in the users' appraisal and perception of something as enjoyable, pleasurable, and fun. As usability goals rely on properties of perceptual affordances, experience goals should rely on hedonomic affordance. Some examples of attributes (see Norman, 2004) that invite positive affect are smiling faces; warmth; comfort; sweet tastes; pleasant smells; bright and highly saturated hues; harmonious music and sounds; rhythmic beats; symmetrical, round, smooth shapes; and an evaluation that indicates a high level of perceived situational control.

Attributes that invite negative affect include darkness; harsh, abrupt sounds; startling noises; information overload; bitter tastes; sharp objects; rotting smells; decaying foods; sudden, bright lights; extreme temperatures; perceived lack of control; and time pressure, among others. Systems that lack functionality and usability invite negative affect in and of themselves (see Bonds & Canak, 1999). Empirical research is certainly needed to link these attributes to experience (Murphy, Stanney, & Hancock, 2003).

Individuation

Individuation (Hancock, 2003) is an individual-centered ethic distinct from human-centered design, which seeks only implementation based on functionality and nomothetic or general human capacities. Individuation is directed to explore ways through which each and every single individual can customize his or her own tools to optimize the pleasure and efficiency of his or her own personal interaction. In this era of burgeoning technology, it is possible to achieve these goals when systems must be customizable and dynamically adaptable to individuals by responding to their affective needs and changing requirements (MacDonald, 2003).

Customization allows for the accomplishment of experience goals on an individual level as users bring their own cognitive appraisals, past experiences, traits, and mood states to the interaction and as these users change their views of the situation and themselves over time. When individuals work in a social context, it is the responsibility of the adaptive technical system to integrate their respective individual preferences for interaction. In hedonomics design, operators are viewed as complex entities with a range of multilevel needs. Therefore, tools must be adaptable to fit the individual needs of the operator, and to do this, we need customization principles. To begin, we suggest two principles: *aesthetic longevity* and *seamless interaction.*

Aesthetic preference is related to a balance between typicality and novelty (Hekkert, Snelders, & Wieringen 2003). The user seeks products that are similar in style to the norm yet are also fresh and creative. In time, however, new ideas that were once popular because they were original become the norm, causing the product to tip toward typicality and lose its novelty. Aesthetic longevity is a customization principle of design that keeps a classic form combined with innovative option features that can be changed over time (for example, face plates on cellular phones). Traditionally, the approach to the study of aesthetics has focused largely on investigating the consensus of what the population finds transiently pleasing. Hedomonics takes an individual-centered approach through customization in recognizing that aesthetics, because of its subjective nature, is essentially idiosyncratic. By incorporating this principle into the development of the product, we seek to create systems that adapt to cultural norms and personal change, hence maintaining the balance between typicality and novelty over time. Aesthetic longevity facilitates a continuous state of newness that consistently elicits pleasurable feeling.

> To measure pleasure, one needs a combination of both subjective and unobtrusive psychophysiological measures to capture its various dimensions.

The principle of *seamless interaction* enables the user to interact optimally with the tool at hand. This in turn facilitates the transparency of the tool, enabling the user to focus effort on task completion and not on the tool itself. In this way, the tool becomes an unconscious extension of the user and is integral to the conscious experience of self (see Hancock, 1997). This promotes the experience of *flow* or being in the *zone of optimal function* as the seamless interaction between the user and the system itself induces intrinsic pleasure. The act of customization should in itself be a seamless interaction enabling

the user to easily adapt the tool to him or herself; but ideally the tool should adapt to the user, the former being passive customization and the latter being active customization.

Through principles such as aesthetic longevity, seamless interaction, and customization, the user's perception of control over the environment is progressively enhanced. Perceived control has been shown to improve job satisfaction, work motivation, and positive affect because it fulfills the need for autonomy (Fritzsche & Parrish, 2005). The function of system stays, but the features can be customized. True individuation would be in a sense unlimited customization, which we see as a feasible possibility in the foreseeable future.

Design Recommendations

As is evident from the foregoing discussion, hedonomics is in its infancy. It may appear premature, therefore, to advance design guidelines for an area of such relative immaturity. However, drawing on allied areas of research, we think we can posit the following initial recommendations:

- Designing for pleasure should be an explicit goal from the initiation of each project.
- Until we understand more of hedonomic design, pleasure as a design goal should remain subordinate to safety, functionality, and usability.
- To achieve hedonomic goals, we need to establish a valid and reliable "measure of pleasure."
- An explicit requirement of hedonomic design is that it is theoretically driven and empirically grounded.
- Hedonomic design achieves its explicit goal with full individuation, which, in combination with general human-centered design, expressly requires a personalized design for each and every individual. This may be achieved by incorporating principles such as aesthetic longevity, seamless interaction, and others that we anticipate will emerge from advances in hedonomics inquiry we have discussed.

Summary and Conclusions

Throughout human history, design has sought to serve the goals of its progenitors. Since humans first fashioned tools, such goals have ranged from the laudable to the lethal. History shows that for much of civilization, the pain associated with tool use has been the lot of the masses, whereas the pleasure that the blessings of technology render has predominantly been the privilege of the few. Some would argue that this imbalance persists in the world today (Moore, 2001).

Fortunately, contemporary technical systems provide the flexibility to achieve both their proximal goal of production while also permitting and promoting pleasurable interaction. This opportunity can change the aversive character of work into an activity that is much more interesting and engaging. Although work and play remain, in most people's minds, fundamentally different conceptions, we can bring some of the elements that make play enjoyable into work. If we are able to achieve this, the invidious division between obligatory work and discretionary play could potentially begin to dissolve. Marxian divisions between different forms of class predicated on the nature of work would no longer be tenable, and technology would further fulfill its true purpose in rendering to all the maximum opportunity for self-realization. We have suggested that this be defined by the term *hedonomics* following Hume's assertion that giving a concept a name makes it a stronger reality (Hume, 1739).

Integrating adaptability into designs and providing users with control each promotes the opportunity for pleasurable interaction.

In the genesis of hedonomics, it is important that the development of the science be theoretically grounded and empirically driven. Pleasure, similar to other subjective constructs like mental workload, stress, fatigue, and situational awareness, is a multifaceted concept. To measure pleasure, one needs a combination of both subjective and unobtrusive psychophysiological measures to capture its various dimensions. Further research is needed in developing an operational definition and assessment of pleasure. Pairwise comparison of affective states may be a pragmatic approach to measurement, but it provides no absolute basis for design perfection. For the science to progress, especially in the affective realm, we have to look to the brain. Fortunately, an emerging confluence between ergonomics and neuroscience may provide a guide: *neuroergonomics*. The fusion of neuroscience with ergonomic application has been proposed by Parasuraman (2003), who notes that all interactions, being brain-mediated anyway, can provide further windows into brain function (also see Hancock & Szalma, 2003).

Hedonomics requires that we investigate more carefully those affective facets of response and the neuropsychological understanding that underpins their occurrence. Integrating adaptability into designs and providing users with control each promotes the opportunity for pleasurable interaction. Although pleasure itself has rarely been the primary goal in work situations, it is often just that in leisure activities. Hedonomics does not seek *dolce far niente* (life without care), as challenge and exploration are two primary sources of pleasure. Rather, it seeks to make technology "convivial" (Illich, 1973) in its opportunity to self-select one's own interactive mode among which pleasure and happiness are prime concerns.

With this change, the *erg* of *ergonomics* need no longer be aversive, "TGIF" (thank goodness it's Friday) will be a thing of the past, and humans will embrace their occupations as both a productive *and* self-fulfilling activity.

REFERENCES

Billings, C. E. (1991). *Human-centered aircraft automation: A concept and guidelines*. Moffett Field, CA: NASA, Ames Research Center.

Blythe, M. A., Overbeeke, K., Monk, A. F., & Wright, P. C. (2003) *Funology: From usability to enjoyment*. Boston: Kluwer.

Bonds, C., & Camak, M. (1999). Your call is important to us...please hold. *Ergonomics in Design, 7(4)*, 9–15.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper.

Flach, J. M., & Dominguez, C. O. (1995). Use-centered design: Integrating the user, instrument, and goal. *Ergonomics in Design, 3(3)*, 19–24.

Fritzsche, B. A., & Parrish, T. J. (2005). Theories and research on job satisfaction. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 180–202). New York: Wiley.

Han, S. H., & Hong, S. W. (2003). A systematic approach for coupling user satisfaction with product design. *Ergonomics, 46*, 1441–1461.

Hancock, P. A. (1997). *Essays on the future of human-machine systems*. Eden Prairie: MN: BANTA Information Services Group.

Hancock, P. A. (1999). Life, liberty, and the design of happiness. In M. Scerbo & M. Mouloua (Eds.), *Automation, technology, and human performance: Current research and trends* (pp. 42–47). Mahwah: NJ: Erlbaum.

Hancock, P. A. (2003, October). *Individuation: Not merely human-centered but person-specific design*. Paper presented at the 47th Annual Meeting of the Human Factors and Ergonomics Society, Denver, Colorado.

Hancock, P. A., & Szalma, J. L. (2003). Operator stress and display design. *Ergonomics in Design, 11(2)*, 13–18.

Hekkert, P., Snelder, D., & van Wieringen, P. (2003). "Most advanced, yet acceptable": Typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology, 94*, 111–124.

Helander, M. G., & Tham, M. P. (2003) Hedonomics: Affective human factors design. *Ergonomics, 46*, 1269–1272.

Helander, M. G., Khalid, M., & Tham, M. P. (Eds.). (2001). *Affective human factors design*. London: Asean Academic Press.

Hume, D. (1739). *An enquiry concerning human understanding: A critical edition*. Oxford: Clarendon Press.

Illich, I. (1973). *Tools for conviviality*. New York: Harper & Row.

Jastrzebowski, W. (1857). *An outline of ergonomics or the science of work based upon the truths drawn from the science of nature*. Originally published in *Nature and Industry*, No. 29 and No. 30 (1857); reprinted by Central Institute for Labour Protection, Warsaw, 1997.

Jordan, P. W. (2000). The four pleasures: A framework for pleasure in design. In P. W. Jordan (Ed.), *Proceedings of the conference on pleasure based human factors design*. Groningen, Netherlands: Phillips Design.

Khalid, H. M. (2004). Conceptualizing affective human factors design. *Theoretical Issues in Ergonomics Science, 5(1)*, 1–3.

Karwowski, W. (2000). Symvatology: The science of an artifact-human compatibility. *Theoretical Issues in Ergonomics Science, 1*, 76–91.

Krippendorff, K. (2004). Intrinsic motivation and human-centered design. Theoretical Issues in *Ergonomics Science, 5(1)*, 43–71.

Lazarus, R. S. (2003). Does the positive psychology movement have legs? *Psychological Inquiry, 14(2)*, 93–109.

Liu, Y. (2003) The aesthetic and the ethic dimensions of human factors and design. *Ergonomics, 46*, 1293–1305.

MacDonald, A.S. (2003). Affective technology and affective human factors design in an era of rapid technological change. In *Proceedings of the XVth Triennial Congress of the International Ergonomics Association. IEA, Seoul, Korea.*

Marek, T., & Pokorski, J. (2004). Quo vadis, Ergonomia? – 25 years on. *Ergonomia, 26(1)*, 13–18.

Maslow, A. H. (1970). *Motivation and personality* (2nd ed.). New York: Viking.

Moore, M. (2001). *Stupid white men*. New York: Regan Books.

Moray, N. (1993). Technosophy and humane factors: A personal view. *Ergonomics in Design, 1(4)*, 33–39.

Murphy, L. L., Stanney K., & Hancock, P. A. (2003). The effect of affect: The hedonomic evaluation of human-computer interaction. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 764–767). Santa Monica, CA: Human Factors and Ergonomics Society.

Mynatt, E. D., Melenhorst, A. S., Fisk, A. D., & Rogers, W. A. (2004). Aware technologies for aging in place: Understanding user needs and attitudes. *IEEE Transactions on Pervasive Computing, 3*(2), 36–41.

Norman, D. A. (2004). *Emotional design: Why do we love (or hate) everyday things*. New York: Basic Books.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science, 4*(1–2), 5–20.

Pepperberg, I. M. (1999). *The Alex studies: Cognitive and communicative abilities of grey parrots*. Cambridge, MA: Harvard University Press.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Oxford: Oxford University Press.

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Beyond human-computer interaction*. New York: Wiley.

Seligman, M. E. P., & Csiksentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*(1), 5–14.

Snyder, C. R., & Lopez, S. J. (Eds.). (2002). *Handbook of positive psychology*. Oxford: Oxford University Press.

Suedfeld, P. (2001). Applying positive psychology in the study of extreme environments. *Human Performance in Extreme Environments, 6*(1), 21–25.

Teague, R. C., & Whitney, H. X. (2002). What's love got to do with it: Why emotions and aspirations matter in person-centered design. *User Experience, 1*(3), 6–13.

Vicente, K. (2004). *The human factor: Revolutionizing the way people live with technology*. New York: Routledge.

Wise, J. A., & Fey, D. (1981). Principles of centaurian design. In *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 245–253). Santa Monica, CA: Human Factors and Ergonomics Society.

Woodruff, P. (2001). *Reverence: Renewing a forgotten virtue*. Oxford: Oxford University Press.

This work was conceived and begun in 1997 while the first author was deputy director of the Liberty Mutual Research Center for Safety and Health in Hopkinton, MA. Hancock gratefully acknowledges the support of Liberty Mutual and a number of other agencies for sustained support, which fostered the general conceptions conveyed in this article. The MURI Grant from the U.S. Army and the direction and guidance of Elmar Schmeisser, Sherry Tove, and Mike Drillings has been especially influential. The views expressed here are those of the authors and should not necessarily be construed as representative of the position of any of the agencies acknowledged. ED

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Transportation Research Part F 8 (2005) 1–18
TRANSPORTATION RESEARCH PART F
www.elsevier.com/locate/trf

Police officers seat belt use while on duty

T. Oron-Gilad [a,b,*], J.L. Szalma [a,b], S.C. Stafford [a], P.A. Hancock [a,b]

[a] *Department of Psychology, University of Central Florida, P.O. Box 161390, Orlando, FL 32816-1390, USA*
[b] *Institute for Simulation and Training, University of Central Florida, P.O. Box 161390, Orlando, FL 32816-1390, USA*

Received 23 April 2004; received in revised form 6 October 2004; accepted 28 October 2004

Abstract

Typical seatbelt designs can interfere with police officers' operational work by lengthening their response time in threatening situations. Therefore, in certain operational circumstances there is a direct conflict between *operational* safety (effective response to threat) and *driving safety* (seatbelt use). To evaluate this potential conflict, 341 police officers from the southeastern US completed a questionnaire that included work related and non-work related seatbelt usage information. Factor analysis revealed five influential and significant factors; (1) travel context, (2) crime context, (3) confidence in seatbelt design, (4) speed and distance of travel, and (5) seatbelt ergonomics. These results confirm that seatbelts themselves in police cruisers currently represent a real safety concern of police officers in high threat circumstances.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: Seat belt use; Police officers

I always wear my seatbelt when off duty. I have found that while on duty (for me) it prevents me from exiting the car quickly—we don't know what or when something bad is going to happen—but when it does I won't be strapped in my car dead. If there was a better mechanism for the seatbelt—I would use them on duty (Anonymous Police Officer).

* Corresponding author. Address: Department of Psychology, University of Central Florida, P.O. Box 161390, Orlando, FL 32816-1390, USA. Tel.: +1 407 823 0923; fax: +1 407 823 0921.
E-mail address: torongil@mail.ucf.edu (T. Oron-Gilad).

1369-8478/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trf.2004.10.005

1. Introduction

There have, and continue to be significant social efforts to increase seatbelt usage since for most vehicle users, increased compliance means increased safety in travel. However, for police officers working in high crime areas this is not necessarily true and there exists the real potential for a tradeoff between driving (transport) and operational (task) safety. Indeed, patrol duty is one of the most dangerous assignments for police officers (Barker, 1999). "*Patrol officers are the front-line officers—the most likely to be injured or killed, the first on any crime scene and, usually, the first representative of the department encountered by victims, suspects, and witnesses*" (Barker, 1999, p. 46). Despite the dangers involved, patrol is still viewed by officers as crucial and indeed the core of 'real' police work (Skolnick, 1966). Patrol officers generally work in three shifts; the day shift, the middle (afternoon/early evening) shift, and the night shift. There are distinctive differences in call types and workloads associated with different shifts (Barker, 1999). Day shift corresponds most with the normal working hours, while mid-shift finds people leaving work with more calls on business robberies, gang activities, drug and drinking problems, and more domestic violence. During the later part of the night shift it is easier to separate '*good guys from the bad*' (Barker, 1999, p. 59) since most law-abiding citizens are asleep or at least remain in their homes. During this shift especially, officers tend to inform each other of their location and if possible, join in backup of their fellow officers. According to the New Jersey State Police most of the assaults on officers occur between 6 PM and 4 AM (the middle and night shift), with a peak around midnight. The lowest number of assaults is during the morning hours (4 AM to 2 PM) and it starts to increase after 2 PM as shown in Fig. 1. It is therefore not surprising to expect difference in officers' perception of risk and that such perception of threat varies across work shifts.

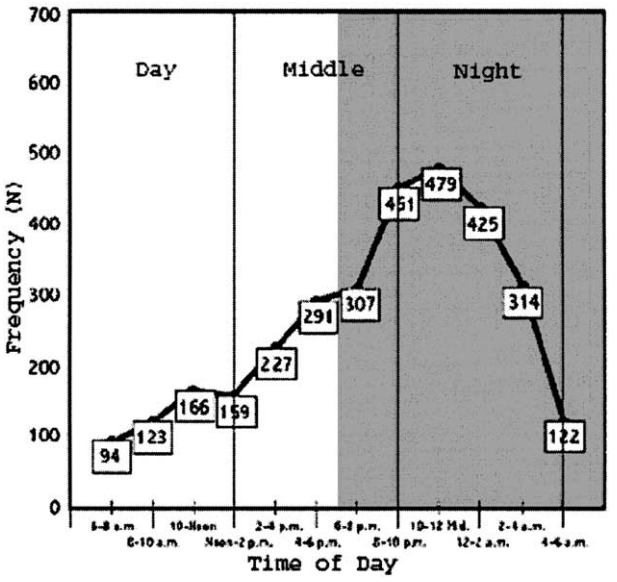


Fig. 1. Assaults on police officers by hour of day, in state of New Jersey, USA, in the year 2000 (New Jersey State Police Official Website, 2004).

The patrol officer is often the first responder among the various emergency services which include police, fire, and paramedic units. Indeed, patrol officers spend most of their time responding to calls for assistance. According to a study by Boston Police Department activities, approximately 55% of these calls require immediate or at least medium priority response (Nesbary, 1998). The records of the New Jersey Police show that in the year 2000, 28% of all the police assaults occurred when officers responded to disturbance calls (New Jersey State Police Official Website, accessed 2004). Response time to high priority calls varies among the many law enforcement agencies. The acceptable range within the Southeastern region of the United States is estimated as 2–3 min. This time estimate includes the response of the 911 call taker, communication with dispatch to the police officer and the actual driving time to the scene. The speed and safety of the patrol officer arriving at a call for help depends on the officers skills and the technology on hand. Patrol officers' work relies mostly on the ''police squad car'' which they use to get to a destination quickly and safely. The technology of the police squad car has developed in recent years. Cars are now commonly equipped with laptops which can access data bases of criminal records for identification matching and background checks. Squad cars have become ruggedized to handle the stress of high performance and high demand driving. Special strobe-light bars have been developed which offer a greater degree of visibility for officers operating in an emergency mode. The improvements of the squad car are mainly technology-driven and done mostly by adding systems to a standard vehicle. At the same time, safety devices, such as the seat belt, remained unaltered and their standard features have not been modified in respect of the added equipment or in accordance with other officers needs. For instance, patrol officers are required to enter and exit their vehicle much more often than ordinary travelers. This wears out seatbelt mechanisms that were not built for such an exhaustive usage. In recent years, populations with special seatbelt needs have been identified; among those are the elderly and the very young children. However, patrol officers seem not to be considered as a population with special needs, possibly due to the sensitivity of the issue—no police officers would like to admit openly that they are not using their seatbelt—and also possibly due to the fact that this issue was never brought back to the car manufacturers as a design requirement.

Officers on patrol duty are expected to comply with seatbelt laws and serve as an example for the citizenry. Seatbelt usage has been estimated to be 40–50% effective in preventing fatalities (Cummings, Wells, & Rivara, 2003; Partyka, 1988). Possibly, the high exposure of officers to traffic accidents and their effects makes officers more aware of the importance of safety devices such as seatbelts than any other population (see Groeger, 2000; Groeger & Chapman, 1996 on the perception of danger in driving). Yet, there has been very little research regarding police officers driving habits, particularly seatbelt use. In regard to driving ability, Dorn and Brown (2003) have found that police officers do not perceive themselves as better drivers than others (see Groeger & Grande, 1996 on driver's self assessment of skill). On the contrary, officers claim that they need more specific driving training (e.g., night driving training). Dorn and Brown's findings are very important since they contradict claims that police officers perceive themselves as better drivers or have more confidence in their own driving ability (an observation which has been found among young male drivers in general e.g., Groeger & Brown, 1989). Again, this is possibly due to the officers' high exposure to traffic accidents.

It is the case that the typical/standard seatbelt design can interfere with police officers' operational work, for example, by lengthening response time to a threatening situation. Thus, in certain

operational circumstances there is a direct conflict between *operational* safety (effective response to the proximal threat) and *driving* safety (seatbelt use). In the past two years, shootings and automobile accidents were cited as the leading causes of death among police officers (FBI National Press Office, 2003; National law memorial fund, 2004). According to preliminary numbers recently released jointly by the National Law Enforcement Officers Memorial Fund (NLEOMF) and the Concerns of Police Survivors (COPS) 82 law enforcement officers were killed in the United States in the line of duty during the first six months of 2004. This represents more than a 7% increase from the 76 officers who lost their lives during the same period in 2003. Over the last 10 years, an average of 166 officers have died annually. Of the 82 officers, 33 were shot, 24 died in automobile accidents, nine officers succumbed to job-related illnesses, seven were struck by automobiles while outside of their own vehicles, four died in motorcycle accidents, two officers were killed in an aircraft accident, one officer was beaten to death, one officer drowned, and one officer fell to his death. Within this delicate balance between operational and driving safety there are many specific situations where the use of the seatbelt not only does not enhance the police driver's safety but actually creates a higher risk by preventing quick maneuvers and the easy access of the officer to his/her weapon. As a consequence, we need to know much more about these specific situations and tradeoffs in order to seek and apply revised designs that permit optimal safety for officers in respect to both sources of threat. The primary goal of the present study was to examine this conflict by providing information concerning officers responses to these two sources of safety concerns.

2. Method

2.1. Participants

The questionnaire was distributed to 600 members of Police Departments in municipalities in the southeastern United States. More than half, 56.8% (341 of 600) returned the questionnaire. These officers volunteered to participate in the study without compensation for their time or effort.

2.2. Questionnaire and Measures

A pilot study was conducted prior the present study. In this initial procedure the questionnaire was distributed to 20 police officers who completed the survey and provided feedback as to the appropriateness of the individual questions. In order for officers to freely discuss their ideas and opinions during feedback sessions they were assured of confidentiality and anonymity. The police officers were informed that none of the information collected would be identified with individuals and that the data would not be compiled and tabulated by either squad shift, or by police department. It is important to note that the police officers specifically asked for a 'free text' section to provide written feedback on the questionnaire so that they could express their own individual ideas.

The final questionnaire consisted of 62 items divided into six sections: (1) demographic information (9 items); (2) personal habits regarding seatbelt use at home and at work (28 items) on a

scale of 1–5 (1-never and 5-always) ; (3) estimations of number of daily events related to seatbelt usage (3 items); (4) opinions regarding seat belt use during a work shift (15 items) on a scale of 1–5 (1-strongly disagree and 5-strongly agree); (5) seatbelt wearing and releasing time related issues (6 items) on a scale of 1–3 (1-satisfactory and 3-needs a lot of improvement); and (6) a free text area for officers to express their own opinion. The questionnaire included a cover letter describing the general purpose of the study and assuring the anonymity of the responses. The cover letter also clearly indicated that the filling of the questionnaire was completely voluntary and that they have to participate in the study if they did not wish to.

2.3. Procedure

Questionnaires were administered to officers at the beginning of their shift by their shift commander. Officers were instructed to complete the questionnaire by the end of their shift and to deposit it in specified location (mailbox) to be collected later by the experimenter. Officers *did not* return the questionnaire to their shift commander. The questionnaire was distributed to police officers with great care as to ensure anonymity and confidentiality. It should be emphasized that the experimenters went to great lengths to provide assurance to the police officers that the data collected would not be individually identified or used in relation to the police officers or the police department in any fashion. To assure confidentiality, we sacrifice analysis by both department, and region. The officers were told that their individual data would be collapsed with other officer's data from various police departments. The sensitive nature of the topic had all those who were involved in the experiment constantly working to ensure the confidentiality.

3. Results

The effects of work shift, perceived threat, and emergency on seatbelt usage were analyzed. In addition, free response feedback, which was considered very important by the police officers who participated in the first stage of the study was also analyzed.

3.1. Socio-demographic and shift work characteristics

Table 1 summarizes the socio-demographics of the 337 police officers who were included in the analysis. Their average age was 37 years (range 21–57 years). Their average years of service were 11 years (range 0–32). Handedness and Body Mass Index (BMI) were also recorded in order to see whether physical dimensions or handedness had an effect on seatbelt usage. Four questionnaires (out of the 341 received) were excluded from the analysis due to a very low rate of questions answered.

The departments included in this survey operated in fixed shift assignments which means that officers were assigned to a shift permanently (or until they request a transfer). Three shifts were examined: day (starting from 5 to 9 AM and ending between 3 and 6 PM), mid-shift (starting from 12 to 2 PM and ending between 8 and 11 PM), and night shift (starting from 6 to 10 PM and ending between 6 AM and 9 AM the next morning). The distribution of officers according

Table 1
Distribution of participants by socio-demographic variables

Variables	Mean	SD
Gender	N	
Men	289 (86%)	
Women	48 (14%)	
Age	37	8
Rank	N	
Officer	258 (77%)	
Corporal and higher	79 (23%)	
Years of service	11	8
Body mass index (*BMI*)	28.2	4.8
Handedness	N	
Left handed	24 (7%)	
Right handed	311 (93%)	
Missing data	2 (0%)	

to their assigned shift was: 49.7% assigned to day shift, 20.7% assigned to mid shift and 29.6% to night shift.

3.2. Factor analysis

An oblique factor analysis was used in one initial run for 43 items; the personal habits regarding seatbelt use at home and work (28 items), and the opinions regarding seatbelt use during a work shift (15 items). The final factor solution for the analysis met the following criteria: (a) each one was based on factors with eigenvalue > 1.0; (b) individual items correlated with the factor concerned at the 0.60 level or above; and (c) only items with a communality of >.50 were selected. Due to missing cells only 235 (70% of the total) participants were included in the factor analysis. The distribution of these officers according to their assigned shift was: 119 (51%) assigned to day shift, 48 (20%) assigned to mid shift and 67 (29%) to night shift, and 1 missing data (0%).

Five factors emerged that influenced the use of seatbelts as shown in Tables 2 and 3: (1) work-related travel (travel context); (2) seatbelt use at varying threat levels (crime context) (3) confidence in seatbelt at varying threat levels (confidence in seatbelt design); (4) speed and distance

Table 2
Factor correlation matrix

Factor	Travel context (travel tasks in police work)	Crime context (threat level)	Seat belt context (confidence in seat belt design)	Speed and distance of travel	Ergonomics of the seat belt
1	1.00	**−.51***	**−.50***	**.68***	.08
2		1.00	.37*	−.21*	−.02
3			1.00	−.29*	.26*
4				1.00	.09
5					1.00

* Correlation significant at $p > .01$.

Table 3
Analysis of the 45 items of seatbelt usage following the oblimin rotation (only individual items that correlated with a factor at a level 0.6 or above were included)

	1	2	3	4	5
Q1-SB usage and traffic accident	.93			.72	
Q2-SB usage in drug problem area	.66	−.96			
Q3-SB usage in high crime area	.66	−.95			
Q4-SB usage in low crime area	.86	−.63			
Q5-SB got caught in equipment in emergency					.88
Q6-SB got caught in equipment in non-emergency					.90
Q7 -----					
Q8 -----					
Q9 -----					
Q10-SB usage in residential areas	.83	−.71			
Q11-SB usage at 25 mph	.89				
Q12-SB usage at 35 mph	.96			.62	
Q13-SB usage at 45 mph	.97			.66	
Q14-SB usage at 55 mph	.96			.70	
Q15-SB usage at 65 mph	.93			.75	
Q16-SB usage in the rain	.91			.77	
Q17-SB usage while backing up another officer	.84			.82	
Q18-SB usage in low visibility conditions	.91			.77	
Q19-Emergency response within 0.5 mile	.71			.63	
Q20-Emergency response within 1 mile	.78			.75	
Q21-Emergency response within 2 miles	.84			.85	
Q22-Emergency response within 4 miles	.84			.92	
Q23-Non-emergency response within 0.5 mile	.86				
Q24-Non-emergency response within 1 mile	.94			.62	
Q25-Non-emergency response within 2 miles	.97			.68	
Q26-Non-emergency response within 4 miles	.96			.72	
Q27 -----					
Q28 -----					
Q32 -----					
Q33 -----					
Q34-It takes too long take off SB in a hurry			.69		
Q35-Pedestrian approaches in low crime area			−.80		
Q36-Pedestrian approaches in high crime area			−.83		
Q37-Confidence in SB in high crime area			−.87		
Q38-Confidence in SB in low crime areas			−.87		
Q39-Confident in overall performance of SB			−.86		
Q40-I often take my SB off in a high crime area		.73			
Q41 -----					
Q42 -----					
Q43 -----					
Q44 -----					
Q45 -----					
Q46 -----					

--- Marked questions were not contributing to any factor.

of travel, and (5) experience with the seatbelt getting caught on equipment (seatbelt ergonomics). These accounted for 63.9% of the total variance. Labels for the five factors have been derived through examination of the salient factor loadings. The first factor accounted for 47.6% of the variance and consisted of a broad spectrum of items that are related to the driving in police work in various threat and emergency situations. Hence this factor reflects upon the travel demands and the travel context of police work. Possibly, this substantive amount of explained variance reflects the high usage rate of seat-belts among officers and also a high level of awareness to the importance of seatbelt. Considering the fact that officers are very often the first official emergency responders to arrive to any traffic accident, it is reasonable to assume that they are highly aware of the importance of seatbelts (see also Groeger, 2000). The second factor accounted for 5.2% of the variance and consisted of items related seatbelt use under different levels of threat. The third, fourth and fifth factors accounted for 4.5%, 4.0%, and 2.6% of the variance, respectively. The third factor consisted of items related to confidence in seatbelt at different levels of threat and reflects mainly upon design limitations of current seatbelts. The fourth factor was related to two sets of items, one that dealt with travel speed and the other that dealt with distance from destination of call under varying levels of emergency, this factor reflects upon the nature of patrol work—mainly responding to various calls. The fifth factor reflected upon two specific items related to the frequency of the seatbelt getting caught on the officers' equipment in emergency and non-emergency situations.

The second, third, and fourth factors were correlated with the first factor as can be seen in Table 2. The reason we chose to leave these factors in the analysis is that each one of them reflects upon a different dimension related to police work and the usage of seatbelt and by maintaining these variables we were able to gain better understanding of these dimensions. Factors 2 and 3 had a negative correlation with Factor 1, indicating that these two factors reflect upon the major threatening situations where officers choose *not* to wear their seatbelts. Factor 4, on the contrary, is positively correlated with factor 1, again reinforcing the use of seatbelt at any time, especially when traveling a distance, or at high speed regardless of the emergency of the call. Factor 5 is moderately correlated with Factor 3 indicating that they both reflect upon the current seatbelt design.

To determine the effect of shift and rank (officer versus corporal and higher) on each factor, a two-way ANOVA was conducted. For Factor 1 the main effect of shift was significant ($F(2,234) = 9.82$, $p < .0001$, $f = .04$), and the main effect of rank was not ($p > .05$). For Factors 2 and 3 the main effects of both shift and rank were significant ($F(2,234) = 12.54$, $p < .0001$, $f = .07$) and ($F(1,234) = 10.03$, $p < .005$, $f = .06$), ($F(2,234) = 6.28$, $p < .005$, $f = .08$) and ($F(1,234) = 10.28$, $p < .005$, $f = .02$), accordingly. For Factor 4 only the main effect of shift was significant ($F(2,234) = 5.07$, $p < .01$, $f = .04$). A possible explanation for the effect of rank on Factors 2 and 3 is that unlike regular officers, higher-ranking officers are generally not the first to arrive at crime scenes. This reduces their level of perceived risk in high threat environments. Post-hoc comparisons (Tukey-HSD) on shift for Factors 1–3 showed significant differences between the day shift and the other two shifts ($p < .01$) but no significant differences between the middle and the night shift. For Factor 4, there were only significant differences between the day shift and the night shift ($p < .01$). The significant difference between the day and the other two shifts is probably due to the difference in threat level and in the perceived risk of being assaulted. There were no significant effects for shift or rank for Factor 5, which is quite expected

due to the items that this factor reflects upon. The effect of handedness (left handed or right handed) and BMI were also examined on each factor but had no significant effects ($p > .05$).

3.3. The effect of high versus low threat on seatbelt use

Factor 2 reflects the questions that are of particular relevance to the effect of threat level on seatbelt use. Questions 2, 3, and 40 specifically addressed seatbelt usage in high crime areas while Questions 4 and 10 related to seatbelt use in low crime areas. Overall officers reported more usage of seatbelt in low crime areas (Q4 and Q10) than in high crime areas (Q2 and Q3). On a scale of 1–5 [1-never, 5-always], the average scores were 4.06 (SD = 1.25) and 3.85 (SD = 1.35) versus 3.15 (SD = 1.56) and 3.18 (SD = 1.59), respectively. Scores were above the midpoint of the scale for both high and low threat areas, implying that most of the time the officers were using their seatbelts.

To determine the effect of environmental threat and shift on officers' estimated frequency of use of their seatbelt, the following analysis was conducted. Threat was defined as high versus low crime areas (Q3 and Q4). Three shifts were examined: day, middle, and night shifts, as well as the officers' rank (officer versus corporal and higher). A 2 (threat) by 3 (shift) by 2 (rank) mixed ANOVA with repeated measures on the first factor revealed a significant effect for threat ($F(1, 322) = 128.76$, $p < .001$, $f = .29$), with officers reporting greater confidence in use of their seatbelts in low threat compared with high threat areas. A significant effect was also observed for shift ($F(2, 322) = 13.82$, $p < .001$, $f = .08$). Post-hoc tests using the Tukey HSD procedure indicated that officers reported greater confidence in seat belt use during a day shift than in either the middle or night shifts. The means for the latter two shifts did not differ significantly from one another. A significant effect was also observed for rank ($F(1, 322) = 9.30$, $p < .005$, $f = .04$), where the higher rank officers reported a higher level of compliance with the seatbelt laws. The interaction of Threat by Shift was significant ($F(2, 322) = 5.84$, $p < .005$, $f = .03$). The interaction between all three factors was not statistically significant ($p > .05$, $f = .01$). Fig. 2 provides the estimated frequency of seatbelt use for Questions 3 and 4 as a function of shift and rank. Fig. 3 provides the level of agreement with the statement that seatbelt are often taken off in high crime area (an inverted version of Q40).

3.4. Confidence in seatbelt at different threat levels

Factor 3 dealt with the confidence in seatbelt performance which may reflect upon seatbelt design. Confidence in seatbelt design was examined by looking at threat level and situation characteristics. To determine the effect of environmental threat and situation characteristics, as well as, shift on officers' confidence regarding their ability to effectively perform their operational duties while wearing their seatbelt, the following analysis was conducted. Threat was defined as high versus low crime areas (Q35 and Q38 versus Q36 and Q37). Two situations were defined; a person approaching the police car and general patrol work (Q35 and Q36 versus Q38 and Q37). Three shifts were examined: day, middle, and night shifts. A 2 (threat) by 2 (approaching person or in general) by 3 (shift) mixed ANOVA with repeated measures on the first and the second factor revealed a significant effect for threat ($F(1, 327) = 77.43$, $p < .001$, $f = .19$), with officers reporting higher confidence in seatbelts at lower threat levels. A significant effect was also observed for the

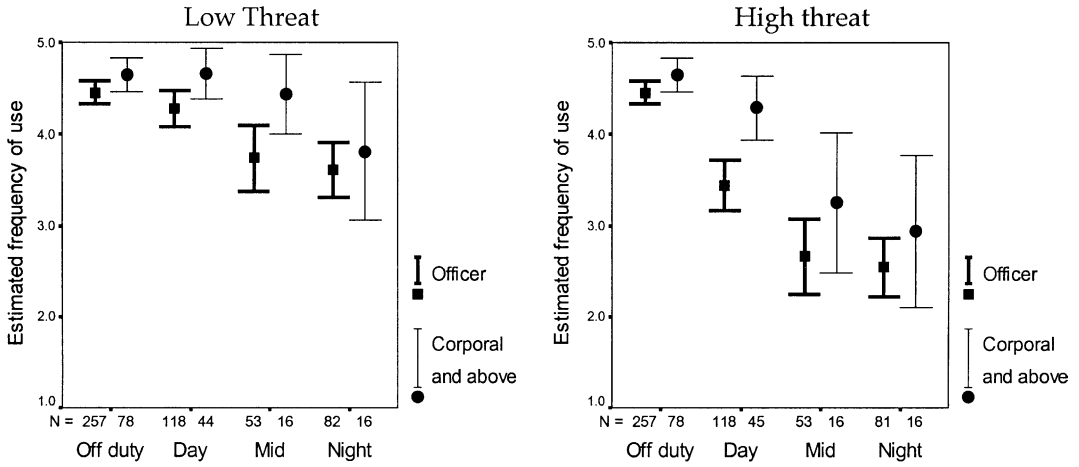


Fig. 2. Estimated frequency of seatbelt use as a function of shift and rank in high and low crime areas (bars represent confidence intervals for mean at 95%).

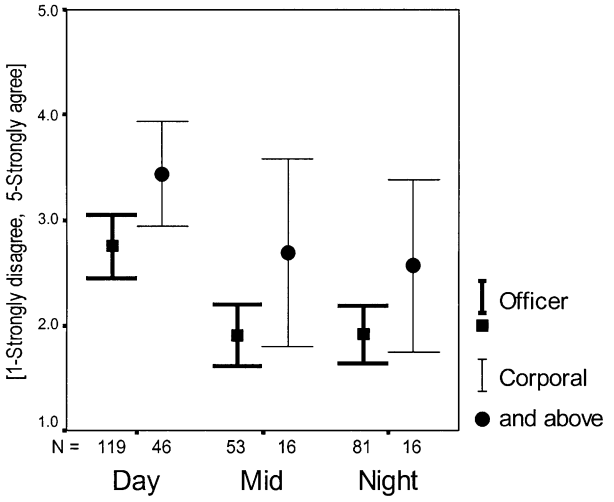


Fig. 3. Level of agreement with the statement "I often leave my seatbelt on in a high crime area" for officers and higher ranks (bars represent 2 standard error of mean).

situation ($F(1,327) = 5.17$, $p < .05$, $f = .02$), where overall, officers had more confidence in the seatbelt during general patrol work than when a pedestrian was approaching their vehicle. A significant effect was also observed for shift ($F(2,327) = 10.02$, $p < .001$, $f = .06$). Post-hoc tests using the Tukey HSD procedure indicated that officers reported greater use of seatbelt during the day shift than in either the mid or night shifts. The means for the latter two shifts did not differ significantly from one another. The interaction for Threat by Situation by Shift was significant ($F(2,327) = 3.23$, $p < .05$, $f = .02$). Fig. 4 provides the estimated confidence in seatbelt as a function of threat, situation, and shift. During the day shift there were significant differences in the low crime area between situations ($t(196) = 3.02$, $p < .005$, $f = .19$), and the confidence in the seatbelt

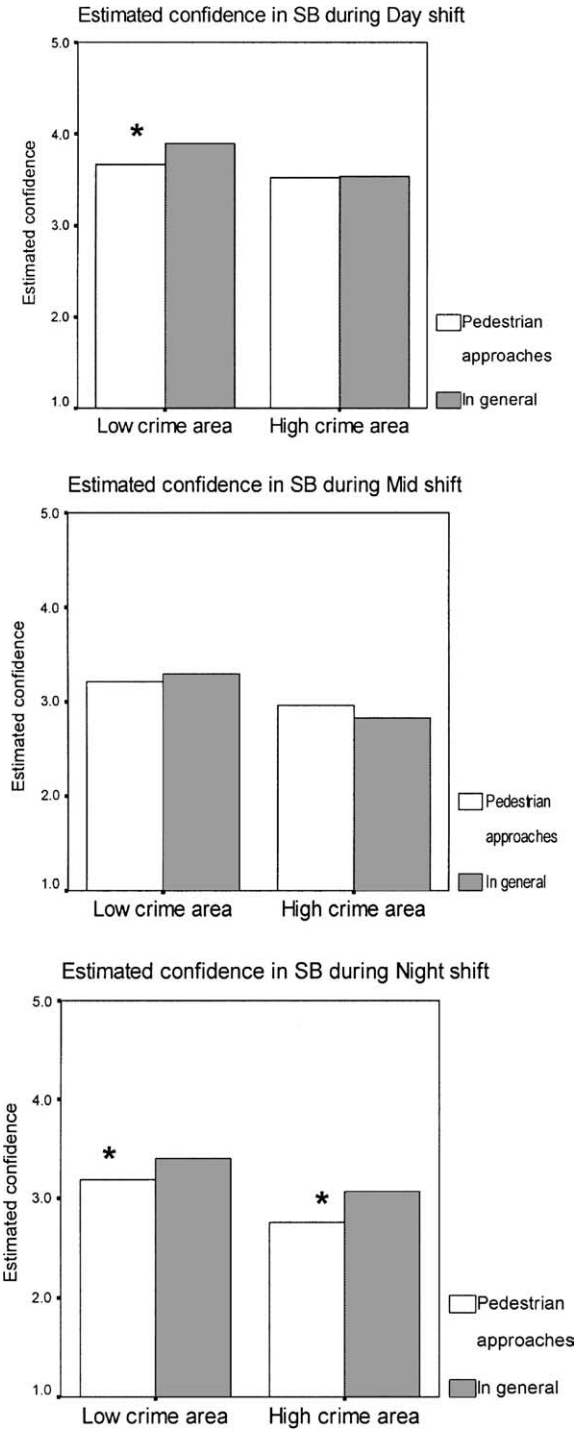


Fig. 4. Estimated confidence in seatbelt design in high and low crime areas as a function the situation and the shift. * Difference significant at $p < .05$, all three main effects were significant.

was lower when a pedestrian approached the vehicle. During mid-shift there were no significant differences between situations. During the night shift we found significant differences between situations for both high and low crime areas accordingly ($t(96) = 2.25$, $p < .05$, $f = .17$ and $t(95) = 2.17$, $p < .05$, $f = .22$), where the estimated confidence in the seatbelt was lower when a pedestrian approached the vehicle.

3.5. The effect of travel speed and duration of drive on seatbelt usage

Factor 4 reflects upon two elements related to responding to calls which are travel speed and the distance from call destination. To determine the effect of travel speed and shift on officers' confidence regarding their ability to effectively perform their operational duties while wearing their seatbelt, the following analysis was conducted. Five speed categories were defined 25, 35, 45, 55, and 65 mph (Q11–Q15). Three shifts were examined: day, middle, and night shifts, as well as the officers' rank (officer versus corporal and above). A 5 (speed) by 3 (shift) by 2 (rank) mixed ANOVA with repeated measures on the first factor revealed a significant effect for travel speed ($F(4, 325) = 63.13$, $p < .001$, $f = .16$), with officers reporting higher use of seatbelt as their travel speed increased. A significant effect was also observed for shift ($F(2, 325) = 8.50$, $p < 001$, $f = .05$). Post-hoc tests using the Tukey-HSD procedure indicated that officers reported greater use of seatbelt during the day shift than in either the mid or night shifts. The means for the latter two shifts did not differ significantly from one another. The interaction between Travel speed by Shift was not statistically significant ($p > .05$, $f = .02$). A significant effect was also observed for rank, $F(1, 325) = 4.57$, $p < .05$, $f = .01$, where the higher ranking officers reported on a higher level of compliance with the seatbelt laws. The interaction between all three factors was also not statistically significant ($p > .05$, $f = .00$). Fig. 5 provides the estimated frequency of seatbelt use as a function of travel speed, shift and rank. The means of the estimated seatbelt use frequencies for all shifts were above the midpoint of the scale (3), indicating that most of the time officers *do* wear their seatbelts while traveling even in the lower speed categories.

In addition, the effect of adverse weather conditions such as rain (Q16) and low visibility (Q18) were examined and in both cases compliance with seatbelt use was high (Mean = 4.57, SD = 3.36 and Mean = 4.31, SD = 1.11, respectively). There were no significant differences between shifts.

3.6. The effect of travel–distance, emergency and non-emergency situations on seatbelt usage

The effect of distance from destination on seatbelt usage was examined by looking at emergency versus non-emergency situations. To determine the effect of distance to the destination, emergency and non-emergency events, and shift on officers' confidence in seat belt, the following analysis was conducted. Four distance categories were defined; half mile, one mile, two miles, and four miles away from the destination (Q19–Q26). Three shifts were examined: day, middle, and night shifts. A 4 (distance) by 2 (emergency and non-emergency) by 3 (shift) mixed ANOVA with repeated measures on the first and the second factor revealed a significant effect for distance ($F(3, 327) = 122.45$, $p < .001$, $f = .27$), with officers reporting higher use of seatbelt as their destination was further away. A significant effect was also observed for emergency ($F(1, 327) = 9.53$, $p < .005$, $f = .03$), where overall seatbelt use was lower in an emergency situation compared with a non-emergency one. A significant effect was also observed for shift ($F(2, 327) = 12.65$, $p < .001$,

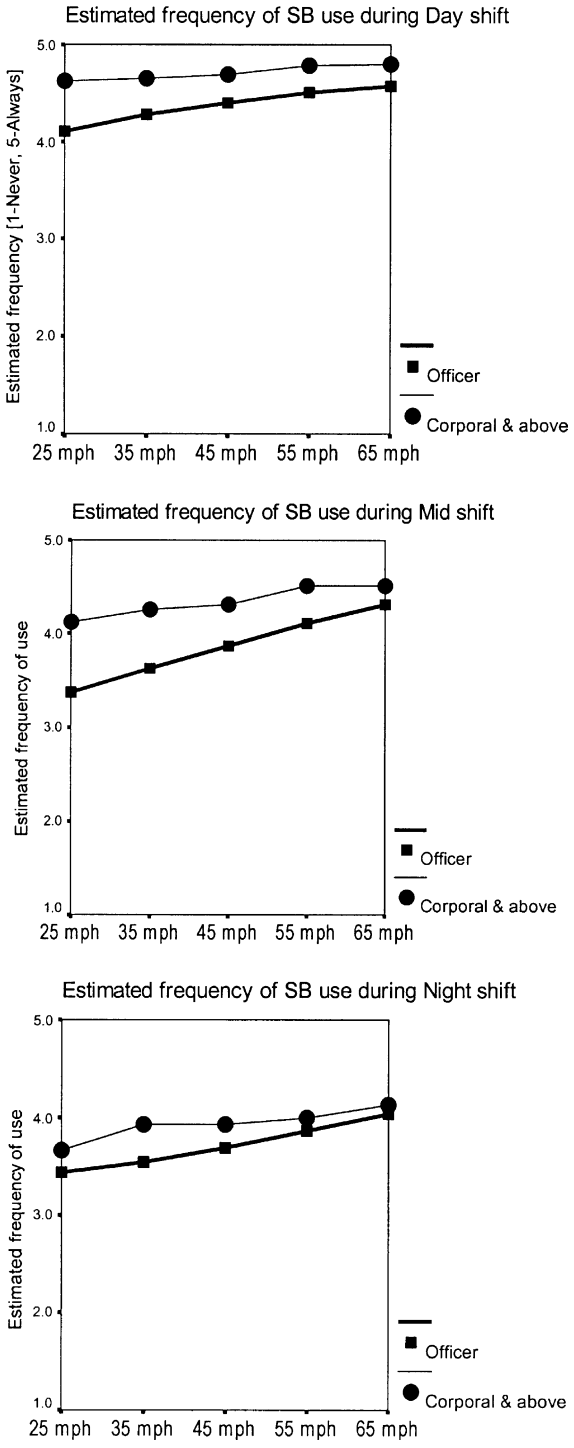


Fig. 5. Estimated frequency of seatbelt use as a function of travel speed, shift and rank.

$f = .07$). Post-hoc tests using the Tukey HSD procedure indicated that officers reported greater use of seatbelt during the day shift than in either the middle or night shifts. The means for the latter two shifts did not differ significantly from one another. The interaction between Distance and Shift was significant ($F(6, 327) = 5.90$, $p < .005$, $f = .04$), and so was the interaction between Distance and Emergency ($F(3, 327) = 24.62$, $p < .001$, $f = .07$). The interaction between all three factors was not statistically significant ($p > .05$, $f = .00$). Fig. 6 provides the estimated frequency of seatbelt use as a function of distance from destination, emergency, and shift. Here again we see that the means of the estimated frequencies of seat belt use among officers were high (above mid point) regardless of the level of emergency or the distance from destination.

3.7. Seatbelt caught on equipment

Factor 5 reflected upon Q5 and Q6 that specifically addressed the question of whether the seatbelt gets caught on equipment while getting out of the car in emergency and non-emergency situations. While there were significant differences between the two ($t(334) = 5.6$, $p < .001$, $f = .19$) the means for both question 5 (mean = 2.85, SD = 1.04) and question 6 (Mean = 2.65, SD = 1.06) were below the midpoint of the scale indicating that such problems did not occur on a regular basis. In Q29 officers were asked to estimate the number of times that the seatbelt got caught in their equipment during an average shift. Forty two (13%) of the officers reported that this never happened to them. One hundred and thirteen (34%) estimated that that seatbelt gets caught 1–2 times per shift. An additional 104 (31%) estimated that it happens 3–5 times per shift, and the remaining 76 (22%) estimated that it happens 6–9 times per shift. This is a pure ergonomic issue and even though it was not perceived by the officers as the main reason for not using their seatbelts, this factor may have contributed to the lack of confidence in seatbelt design reflected upon in Factor 3 since 87% did actually perceive some active interference on a daily basis.

3.8. The effect of training and awareness on seatbelt use

Questions 43–45 addressed the influence of Police Academy training (Q43), field training officers (Q44), and agency policy (Q45) on seatbelt use. A repeated measures ANOVA was significant ($F(2, 334) = 125.9$, $p < .001$, $f = .27$). Indicating that the most influential of the three was the agency policy (mean = 3.47, SD = 1.35), followed by the field training (mean = 2.92, SD = 1.28), and the police academy had the least influence (mean = 2.56, SD = 1.27). These data show that the agency policy regarding seatbelt use is influential on officers' behavior regarding to seatbelt usage.

3.9. Seatbelt use while off duty

Questions 27–28 addressed use of seatbelt while off duty, officers reported on a high compliance level both in their personal vehicle (mean = 4.47, SD = 1.07) and in their police vehicle (mean = 4.50, SD = .97). This is important since it has been suggested that police officers represent just one of a collective group who are personally averse to seatbelt usage (as a baseline for comparison the US national average seat belt use for the year 2003 was estimated at 79% (NHTSA, 2004)). In contrast, this data indicate a high degree of compliance during off duty hours and thus

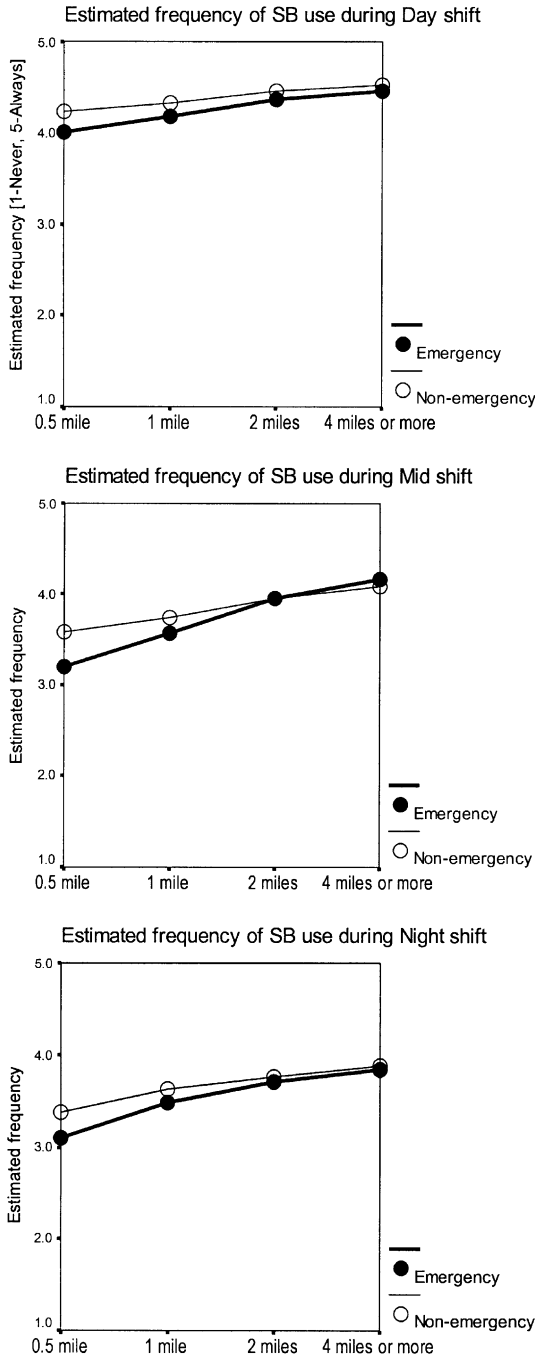


Fig. 6. Estimated frequency of seatbelt use as a function of distance from destination, emergency and shift.

neglect of seatbelt use on duty is not simply a reflection of personal preferences but rather is a rationale response to the specific conditions encountered.

3.10. Required improvements to the current seatbelt settings

The more specific questions (Q47–Q52) dealt with the areas which required improvement in seatbelt design. Here, 51% of the officers thought that the time it takes to *put on* the seatbelt should be improved (Q47). Further, 67% of the officers thought that the time to *take off* the seatbelt should be improved (Q48). Of the total sample, 82% of the officers thought that drawing the weapon while using a seatbelt should be improved (Q49) while only 50% of the officers admitted that they still had a problem in drawing their weapon while not using the seatbelt (Q50). Of the officers, 58% had some problems with the pull mechanism of the seatbelt (Q51). And finally, 74% of the officers would like to see a major change in the seatbelt process.

4. Discussion

For most drivers and occupants of passenger vehicles, the overwhelming concern is for safe passage of travel between origin and destination (see Gibson & Crooks, 1938). That task having been completed, there is in an overwhelming number of cases, the time to remove seatbelts, collect and gather up belongings and exit the vehicle in a simple, orderly manner. In this fashion, they are very much akin to commercial airline passengers who, although they may be in a hurry, find no fundamental conflict between safety and other concerns. If everyday travelers are like airline passengers, police officers are much like Special Forces. They not only have to get to the destination but they must respond quickly and efficiently during that process of arrival otherwise the purpose of their journey may well not be achieved. At the point of arrival, they continue to face Haddon's (1970) kinetic 'tiger' but the source of threat is rapidly evolving from a one-ton vehicle at 60 mph to a one ounce bullet at 1200 mph. The officer cannot afford to ignore either threat at the expense of the other and as the present findings show, they adjust their behavior accordingly.

The collective results support a hypothesis that police officers perceive a conflict between operational and driving safety. These effects are especially evident in high threat situations, and during later work shifts (see Figs. 2–4). The latter results are likely due to the fact that officers typically experience higher workload (e.g., more threatening dispatch calls and more events requiring immediate response) during the middle and night shifts than during the day shift. Our findings further indicate that the current seatbelt designs in police cruisers paradoxically represent a safety problem to the officers, in particular in high crime areas. Although they are aware of the importance of seatbelt use on driving safety (Factors 1 and 4, and also Fig. 5 and Fig. 6). There are situations in which the perceived risk from wearing the seatbelt exceeds the one of not wearing it (Factors 2 and 3). In those high threat situations officers are not concerned with putting their seatbelt on. The question that addressed putting the seatbelt on (Q33) had no significant effect on any of the factors that we identified. It is *taking the seatbelt off* that concerns the officers because the time-to-exit (time to get out of the police squad car) and the time-to-draw (a secured seatbelt can in some cases add another step to drawing a weapon) are perceived as too long with the current vehicle– driver–seatbelt settings. Clearly this indicates that non-usage of seatbelt while on duty is *not* a problem related to officers ignoring of the importance of seatbelt use in vehicles in general. On the contrary, we found that officers use their seatbelt regularly when off duty or when traveling at high speed (even while on duty). Non-usage of seatbelt is a result of a poor,

ill fitting design that endangers the officers in threatening situations because it reduces their mobility and increases the risk of them being superceded by criminals and suspects.

The results of this study imply that the current protective envelope that officers are provided needs to be redesigned and adjusted to their specific needs. In particular there is a need to reduce the time-to-exit and increase officers' maneuverability. For the short and mid-term, we suggest that small changes in the buckling point of the seatbelt be made. Such a change may be able to reduce incidents where equipment gets caught. For the long term a superior internal design of the squad car is required. This design should be use-centered (Flach & Dominguez, 1995) and should include safety solutions that will not interfere with the officers' egress from the vehicle. More suitable seatbelt design may also be useful for other operational situations such as military vehicles (Naylor, 2003) or other work domains that include high-threat environments including peace-keeping operators and emergency service activities.

There is little doubt that seatbelt use saves lives. The higher the degree of compliance, the greater the social protection that this technology renders. However, the belief that 100% compliance to standard seatbelt use is the fundamental goal is to misunderstand the purpose of transportation for some small but select users of the transport system. Certain specific sections of the traveling population have multi-leveled goals, some of which are in a degree of inherent conflict with safe transit. To police officers, this conflict need not necessarily persist but does at present because of the perpetuation of population-based designs which are somewhat of odds with their particular needs. Through our present confirmation of such inherent conflict we seek avenues of reconciliation through design, engineering, and training approaches which will alleviate such conflict and allow those in law enforcement to go about their necessary duties as safely as possible.

Acknowledgments

The authors wish to thank the Heads of the Police Departments who let us perform the survey among their officers. The authors would also like to express their gratitude to the 341 police officers who voluntarily completed the survey and expressed their valuable comments for the sole purpose of contributing to a research study. Karen Delos Santos, our student, who helped in data collection, is also acknowledged. The authors would also like to acknowledge Dr. P. Ward for his contribution to this study. This work was facilitated in part by the Department of Defense Multidisciplinary Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. P.A. Hancock, Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect the policy of any named agency.

References

Barker, J. C. (1999). *Danger, duty, and disillusion: The worldview of Los Angeles police officers*. Prospect Heights, IL: Waveland Press.

Cummings, P., Wells, J. D., & Rivara, F. P. (2003). Estimating seat belt effectiveness using matched-pair cohort methods. *Accident Analysis and Prevention, 35*, 143–149.

Dorn, L., & Brown, B. (2003). Making sense of invulnerability at work: A qualitative study of police drivers. *Safety Science, 41*, 837–859.

Federal Bureau of Investigation National Press Office (2003). *Fact sheet for law enforcement officers killed and assaulted, 2002*. Washington, DC: Federal Bureau of Investigation National Press Office. Available from: http://www.fbi.gov/pressrel/pressrel03/leoka02press.htm.

Flach, J. M., & Dominguez, C. O. (1995). Use-centered design: Integrating the user, instrument, and goal. *Ergonomics in Design, 3*, 19–24.

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile driving. *American Journal of Psychology, 51*, 453–471.

Groeger, J. A. (2000). *Understanding driving: Applying cognitive psychology to a complex everyday task*. Hove, UK: Psychology Press.

Groeger, J. A., & Brown, I. D. (1989). Assessing one's own and others' driving ability: influences of sex, age, and experience. *Accident Analysis and Prevention, 21*(2), 155–168.

Groeger, J. A., & Chapman, P. R. (1996). Judgment of traffic scenes: The role of danger and difficulty. *Applied Cognitive Psychology, 10*, 349–364.

Groeger, J. A., & Grande, G. E. (1996). Self-preserving judgments of skill. *British Journal of Psychology, 87*, 61–79.

Haddon, W. (1970). On the escape of tigers: An ecologic note. *Technology Review, 72*, 44–47.

National law officers memorial fund (2004). *Law enforcement officer fatalities reach 148 for 2003—shootings and automobile accidents cited as leading causes of death*. Press release January 5, 2004. Available from: http://www.nleomf.com/.

National Highway Traffic Safety Administration (2004). Safety Belt use in 2004—overall results. Report number DOT HS B09 783. Available from: http://www.nhtsa.dot.gov/people/injury/Traffic-Safety-Facts/Research-Notes/SeatBelt-Use-2004/images/RN%20Final.pdf.

Naylor, S. D. (2003). Heroes everywhere I went. *Army Times, 63*(44), 8.

New Jersey State Police Official Website (2004). Crime reports and Statistics: 2000 Uniform crime report section X: Law enforcement officers killed and assaulted summary. Available from: http://www.njsp.org/info/ucr2000/sec10ucr2000.htm.

Nesbary, D. K. (1998). Handling emergency calls for service: organizational production of crime statistics. *Policing: An International Journal of Police Strategies and Management, 21*(4), 576–599.

Partyka, S. C. (1988). *Lives saved by seat belts from 1983 through 1987*. DOT HS 807 324. National Highway Traffic Safety Administration, Washington, DC.

Skolnick, J. H. (1966). *Justice without trial*. New York: John Wiley and Sons.

Hancock, P.A., Oron-Gilad, T., & Szalma, J.L. (2005) To Honor Wickens and Re-Stack His Boxes *Paper Applied Attention: From Theory to Practice Conference*

TO HONOR WICKENS AND RE-STACK HIS BOXES[1]

P.A. Hancock, T. Oron-Gilad, and J.L. Szalma[2]

University of Central Florida
3100 Technology Parkway, Suite 337
Orlando, Florida 32826-3281

ABSTRACT

Multiple attentional resource theory represents a construct which served to unify two disparate views of human performance capacity. Subsequently, the theory had a particularly strong influence on the practice of human factors, especially in interface development where it arguably remains the strongest behavioral heuristic for design. This confluence of contribution to both fundamental theory and practical application served to facilitate its pervasiveness and impact of multiple resources. Here, we offer elaborations upon this conception with the fundamental purpose being to honor Wickens and seek to build upon what is evidently, his most significant scientific success.

---

[1] *While this title may seem somewhat irreverent, its purpose is rather exactly the opposite. Of course, the form is not original, it derives from a famous paper by Stevens in Science:*
Stevens, S.S. (1961). To honor Fechner and repeal his law. *Science, 133,* 80-86.

[2] *Correspondence: P.A. Hancock, Department of Psychology and Institute for Simulation and Training,, University of Central Florida, Orlando, FL 32816. Ph:(407)-823-1492/2310. email: phancock@pegasus.cc.ucf.edu website: www.mit.ucf.edu*

TO HONOR WICKENS AND RE-STACK HIS BOXES

PROEM

In the dark and dangerous days of the late 1970's, a wild, young Rableaisian psychologist, (see Figure 1), of doubtful educational provenance (The University of Michigan) but of evident strong genetic background (see Wickens & Wickens, *undated*) battled an issue of bewildering, nay Gordian complexity - human attention. With competing sources of stimulation entering one ear after another (Moray, 1967) and stages and filters shifting uncontrollably across the empirical landscape, psychological research suddenly saw their box-world, dowsed and awash in metaphoric resources (Kahneman, 1973). Where, oh where, is our intellectual Lochinvar to come 'out of the West,' cried the confused psychological masses?

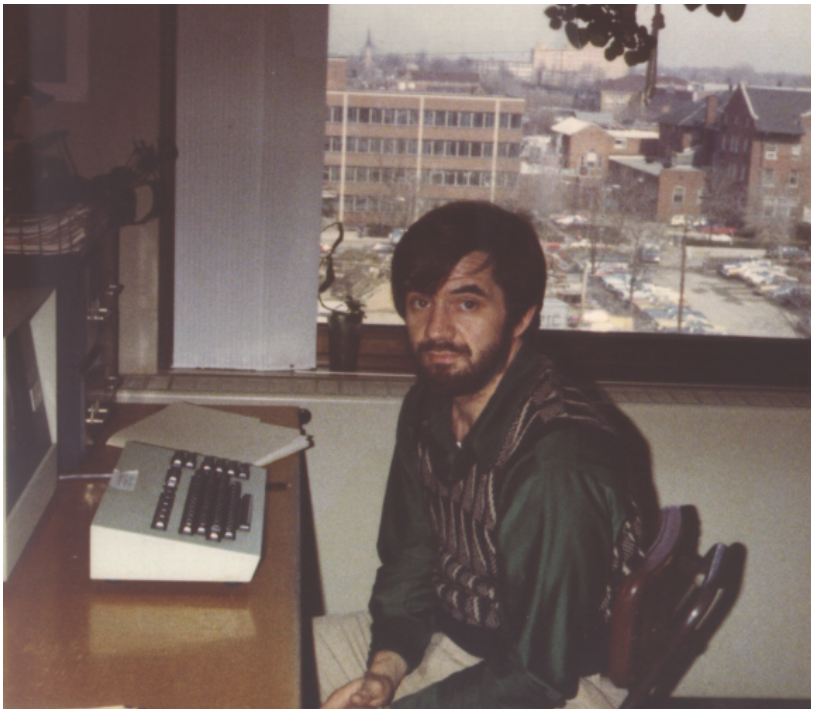


Figure 1: *Is this a computer I see before me? Early days of Champaign and roses.*

Courage they said - and pointed towards the Chris. And, with a veritable Delosian exhibition of brilliance he answered the call! Why don't we just slam the two together, he opined, and thus multiple resources were born[3]. After all, multiple resources give us a greater number of explanatory degrees of freedom and that's never bad for a budding psychological theory (Wickens, 1979; *personal communication*). Attention and attention's laws were hid in darkest night, Pew said let Wickens alone and all was light! And they saw that it explained that pesky *difficulty insensitivity* and pronounced that it was good and gave it several prizes (see Figure 2)! And, for at least a decade (which is, after all, all a decent behavioral theory can really expect) they lived happily ever after. But then…

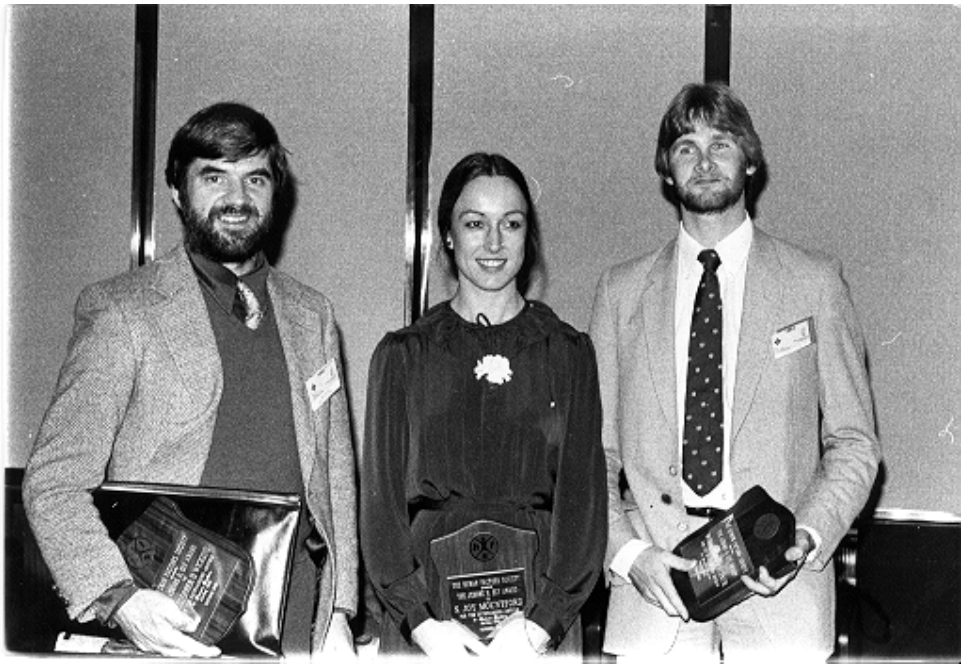


Figure 2: *Christopher Delos Wickens, Shelagh Joy Mountford and William A. Schreiner, the recipients of the Jerome Ely of the Human Factors Society, at the Seattle Meeting of 1982. The authors award winning 1981 publication was entitled:* Multiple resources, task-hemispheric integrity, and individual differences in time sharing, *Human Factors, 23(2), 211-229. It was indeed an arduous and enlightening experiment, after all we should know since one of us was a participant.*

[3] *Let us Passover Navon and Gopher (1979) in silence and never speak of them again!*

Hancock, P.A., Oron-Gilad, T., & Szalma, J.L. (2005) To Honor Wickens and Re-Stack His Boxes *Paper Applied Attention: From Theory to Practice Conference*

TO HONOR WICKENS

Science proceeds in fits and starts. Often there is an original theory that encourages structured experimental investigation and slowly, as the empirical database is assembled, quirks and inconsistencies begin to emerge until there arises a state of almost intolerable dissonance that acts as a goad to theorists to propose new formulations to straddle the behavioral landscape thus revealed. While such new and sweeping insights permit a general brisance of understanding, they can only be accomplished effectively by those steeped in the problem area and intimately familiar with the nuances of the current state of knowledge. In essence, one has to know one's problem in order to generate solutions to one's problem. Wickens' encyclopedic understanding of the attention literature in the late 1970's allowed him to comprehend a vast sweep of issues and so to provide a structured answer which subsequently dominated the literature for many years. The added benefit of his formulation was that it also represented a strong design heuristic which amplified the impact of the theory well beyond the experimental psychological literature alone. However like all theories, it cannot remain static. By introducing differing and extended interpretations of the various axes we hope to indicate some avenues of potential progress through which to enlarge this, the most evident of all Wickens' contributions.

PAY ATTENTION

Single-channel approaches to information-processing are identified largely with the late 1940's and early 1950's when the first blush of information theory penetrated into the psychological sciences (see Shannon & Weaver, 1949). Attention, as an issue, was problematic for these early stage models (see Broadbent, 1958). Traditionally, attention was linked to a filter, a bottleneck, or some form of "rate-limiting" element in the progressive processing sequence but contradictory experimental findings led to much dispute about the nature of the bottleneck and its relative location in the identified sequence (see Moray, 1967; Wachtel, 1967). These issues, while leading to a series of

ingenious experiments, never really reached a definitive conclusion. In contrast to the processing filter approaches, the resource notion (Kahneman, 1973), presented an energetic perspective on attention (and see Freeman, 1948). In this conception, more fluid processing resources can be directed to a task, as compared to the all-or-none allocation that a pristine stage model implies. Following its initial formulation, debate continued over whether there was only one such resource (Kantowitz & Knight, 1976) or several such resources (Navon & Gopher, 1979). In principle, Kantowitz's (1987) notion is fortified by the concept of parsimony. Thus if predictions from a single resource model can be shown to emulate the empirical data satisfactorily, then multiple research models theoretically become redundant (although see Bronowski, 1966 on the fallacy of parsimony). However, of greater importance is Kantowitz's (1987) observation that multiple resource approaches are *too* powerful. That is, they have sufficient degrees of freedom so that almost any pattern of data may be encapsulated by the appropriate manipulation of the variables proposed.

The modal, multiple resource approach is crystallized in Wickens' conception, rendered colloquially as the "box model" (Wickens, 1980). From this conception, which is shown in Figure 1, multiple attentional resources were identified upon the basis of three fundamental dimensions. The first dimension is that of *processing stage*. Fundamentally, this is a temporal axis and reflects the concern for the sequence of encoding, decision-making, and response which were made seriatim (even though the potential for parallel processing was considered). A second identified axis was that of *processing code* consisting of verbal and spatial components of a task. The final axis of *processing modality* was separated into visual and auditory elements. These respective differentiations were made essentially from *post-hoc* interpretations of experimental data, which commonly employed the dual-task technique to produce relevant experimental information. In noticing occasions where two tasks failed to interfere with each other's performance, a prediction which, by the way, is impossible from a strict unitary resource formulation,

Wickens used the concept of *difficulty insensitivity* to distinguish the resource structures given in Figure 3.

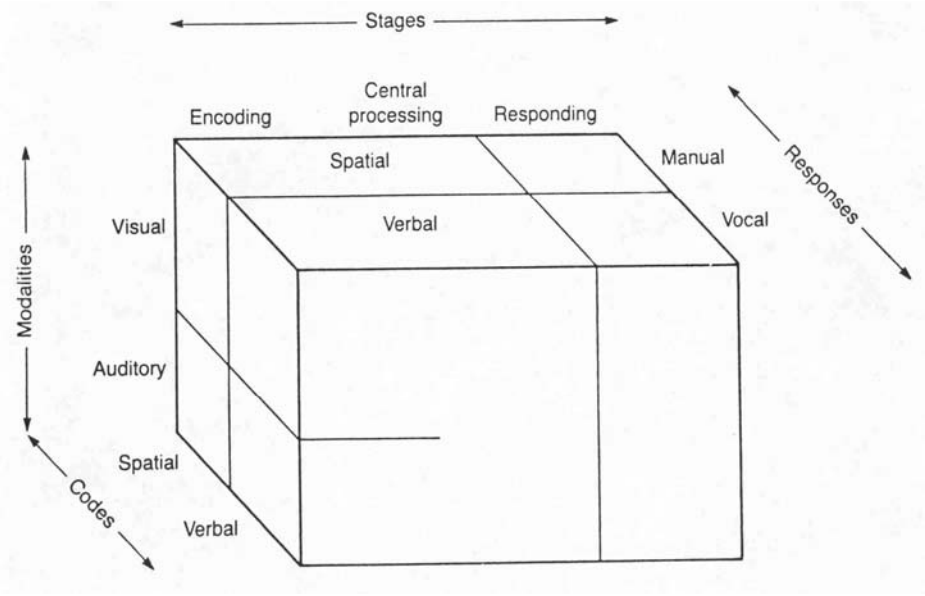


Figure 3: *The classic Wickensian 'Box Model' of human attention and design heuristic for information and response distribution for the human operator (after Wickens, 1980).*

Together with the concept of functional cerebral space (Kinsbourne & Hicks, 1978) and the differentiation of resources as advocated by Navon and Gopher (1979), Wickens was able to define a hybrid model that solved many of the contemporary questions concerning attention. This formulation also proved to have a substantial impact in the realm of human factors and system design. It is not the purpose of the present work to confront the methodological (Damos & Lyall, 1986) or existential (Navon, 1984) attacks that have been directed at the notion of resources in general and multiple resource theory in particular. Rather, the present work examines a number of implications of the model in particular, and of attention and the nature of reality in general.

ROTATING THE BOXES

Processing Codes

The prototypical representation of the box model is given in Figure 3. Examples of representative forms in each of the respective cells of the left-facing end axes have also been given by Wickens and are reproduced here in Figure 4.

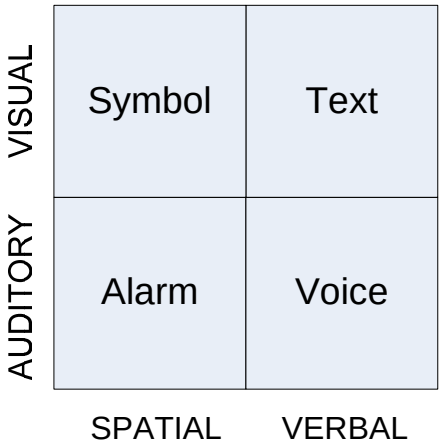


Figure 4: *Examples of members in four end cells in the 'Box Model'. A simple auditory alarm tone in the spatial code is expanded into a voice stimulus in the verbal realm. Similarly, a visual symbol is transmitted to text as the change is made from spatial to verbal code.*

Let us consider the axis of processing code in greater detail. In the original formulation, processing code lies on the base axis and is divided between spatial and verbal components. However, this horizontal differentiation implies a qualitative distinction. But is this qualitative distinction justified as being necessarily so? Perhaps not! In fact, one can argue that any verbal representation can be given as an abstraction of the spatial dimension. Let us consider this in a little more detail. In respect of speech (voice), this is a form of verbal code. However, in reality speech is actually the complex distribution of energy in space-time, and since we shall deal with time as one of the other dimensions, we might be justified in treating verbal representations as special forms of temporarily

enduring spatial codes. Similarly, if the verbal code is expressed in the visual dimension, it is viewed as letters, words, sentences, and the like (or more generically – text). However, these are also each only particular forms of geometric ordering in a spatial configuration and therefore again a particular form of abstraction. Thus, if verbal codes are viewed as particular, high-level abstractions of spatial codes (in reality spatio-temporal codes), then we can rotate the boxes so that this is now graphically represented as an hierarchy of such abstraction. This rotation (only for the left-facing 'end' axes) now appear as shown in Figure 5. Note also that there is a secondary advantage to this rotation. The new base axis is now modality, and while it is true that different sensory systems transduce different ranges and forms of stimulus energy, their parsing into the classic sensory systems is probably more logical since there is no necessary abstraction from vision to audition. The hierarchic abstraction of codes now fits appropriately on the vertical axis.

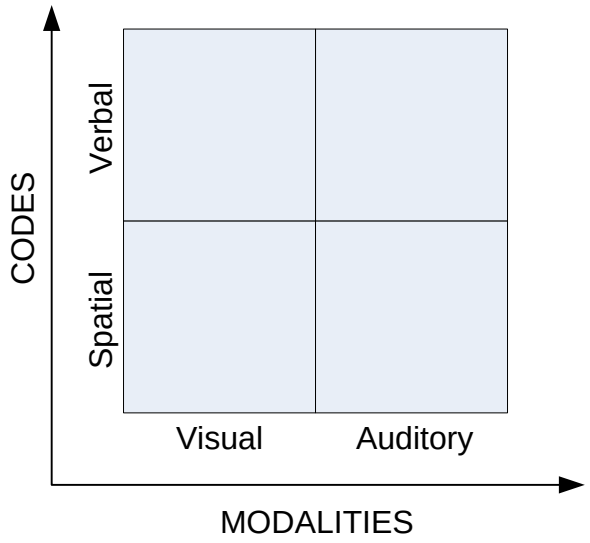


Figure 5: The '*Box Model*' *rotated through 90⁰ in order to emphasize the emergent relationship between the respective codes.*

However, we have not reached the end of the issue of processing code. We suggest that if the verbal level is actually an abstraction of the spatio-temporal level then there are other abstractions beyond the verbal itself. We see at least three other levels. They are i) the informational level, ii) the symbolic level and iii) the meta-abstract level. Perhaps an example will be useful in differentiating these levels. Consider the US *'Declaration of Independence*.' When approaching the document (which is now unfortunately very faded and as a consequence almost illegible), it is evident that there are spatio-temporal graphic markings on the document itself. Further, it is evident also that these graphic images are arranged in the form of (English) words. These observations are evident to *anyone* who sees the document. However, to a non-English speaker, the words themselves, when expressed either in the visual or the aural form are essentially meaningless, as are the tactile-kinesthetically expressed dot patterns of Braille to the uninitiated. In contrast, the individual familiar with English will find a vast amount of information contained in the document. Thus information (which is always a combinational property of the individual and the environment in which they find themselves, and see Hancock, Szalma, & Oron-Gilad, 2005) is a level beyond verbal expression alone. But, this is not all. The actual Declaration is much more than simply another bit of paper. For the people of the United States, and arguably for many in the world beyond American borders, the Declaration has great symbolic significance - *as an object in and of itself*. Indeed, items in the world almost always possess symbolic value as well as spatio-temporal continuity, verbal appellation, and informational content. We believe, however, that we can go one higher level still. This level is the one we have labeled the 'meta-abstract' and is the apex of the abstraction process. Although representative of all of the levels we have been discussing, the Declaration itself goes one step beyond this. It is symbolic *beyond itself as an entity*. Thus, while the physical document itself is of great (monetary) value, its true symbolic value goes well beyond this financial valuation. Thus people can, and do, see in this document intangibles such as freedom, liberty, and hope made manifest. The exact anti-thesis of this meta-level

representation might be seen, for example, in a German document of the Second World War authorizing transportation of Jewish individuals to concentration camps. This latter document would also possess all of the attributes at each abstraction level appropriate to these fluent in German. However, at the meta-level, the attribution is diametrically opposite of the former example. This is crucial to understand, since it means the implications for cognition and subsequently reality, can vary according to interpretations derived at each level[4]. The trick here of course, is to see these abstraction levels in all of the perceivable environment and not simply to restrict this idea to human manufacturanda alone (and see Hancock, 1997). Thus, in the ways we have indicated, and of course potentially in ways we have not, the processing code can go well beyond the simple 'spatial' and 'verbal' differentiation of Wickens' original formulation. Whether there is a differentiated visual attentional resource pool for each of these levels, separated say from an auditory attentional resource pool is indeed a debatable issue. In our view this issue (e.g., can one differentiate between seeing and hearing hope?) is one that remains unresolved. We should note that our elaborations are no criticism of Wickens per se, since as a theorist he could only work with the data at hand. Perhaps future advances will provide such information? The addition of these other code extraction levels now provide us with an adjustment to Figure 5 which is now shown in Figure 6. This representation also implies that not all code levels are necessarily of equivalent importance and that there is a necessary nesting involved.

[4] *Parenthetically, it introduces a very important but complex argument concerning the nature of entropy. For the physicist, the transformation of both document, if they were burned, would be exactly equivalent. As is evident, for the psychologist, for the sociologist, in fact for the everyday individual, the respective acts would be very different indeed.*

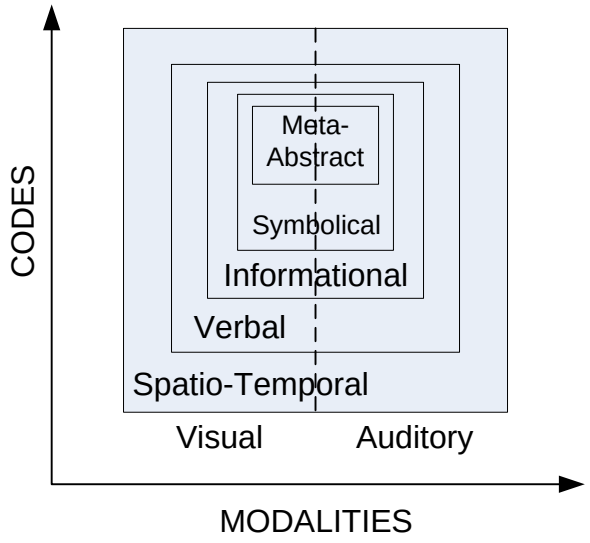


Figure 6: *The Processing Codes revised. The codes are now presented as an abstraction hierarchy.*

One interesting speculation which comes from this hierarchy of abstractions concerns the notion of human language development. While pictographs and orthographic symbology appear to represent our earliest recorded history (e.g., Lascaux, etc), speculation about early human language must perforce, be more tentative. However, it is possible to propose that both picture symbology and, subsequently print, together with spoken language expanded into the functional "space" available. This should certainly be differentiated from the idea of functional cerebral space (e.g., Kinsbourne & Hicks, 1978), but the notion of an ecological niche representing room for expansion is fundamentally relevant to both ideas. It implies that as the state of understanding grows and the pressure to increase the precision of communication increases, another level of the hierarchy is exploited (and/or created?). It is feasible to suggest that human-machine interaction forms the basis for the next level of extraction and that human factors is the language to interpret that emergent abstraction (and see Hancock, 1997).

Processing Modalities

Having considered processing code, let us now consider the new base axis which is thus created namely - processing modality. In identifying the visual and auditory modality, Wickens specified the two major avenues through which any individual may assimilate sensory information. However, this does not exhaust the number of possibilities. As may be obviously extrapolated, the tactile, olfactory, and kinesthetic senses form a series of modalities and while others such as thermal sensation and pain can also be identified it is important to note that the processing modality axis is *not* subject to the potential for an infinite regress but rather is limited to the known array of (predominantly human) sensory transduction systems. In the human, vision dominates (Sivak, 1996). Thus the size of the boxes can be adjusted to reflect the dominance of and contribution of each particular modality. This is illustrated in Figure 7. Clearly for different organisms, the relative contribution of each modality can change. The ascendancy of olfaction in tracker dogs being perhaps a pertinent example (Budiansky, 2000). As we shall eventually seek to conclude that the derived universe of extended Wickens' boxes actually composes the model from which any organism derives their 'reality,' it is important to understand that there are significant cross-species differentiations in these respective 'universes.' That reality also varies between individuals within any single species has also not escaped our attention. More could certainly be said about the ordering, the impact, and the integration of the respective sensory systems and the neuroscience of these capacities would well inform a 'corticotopic' rather than box representation. However, in principle, our case is made and thus, rather then belabor the point we pass on to processing sequence.

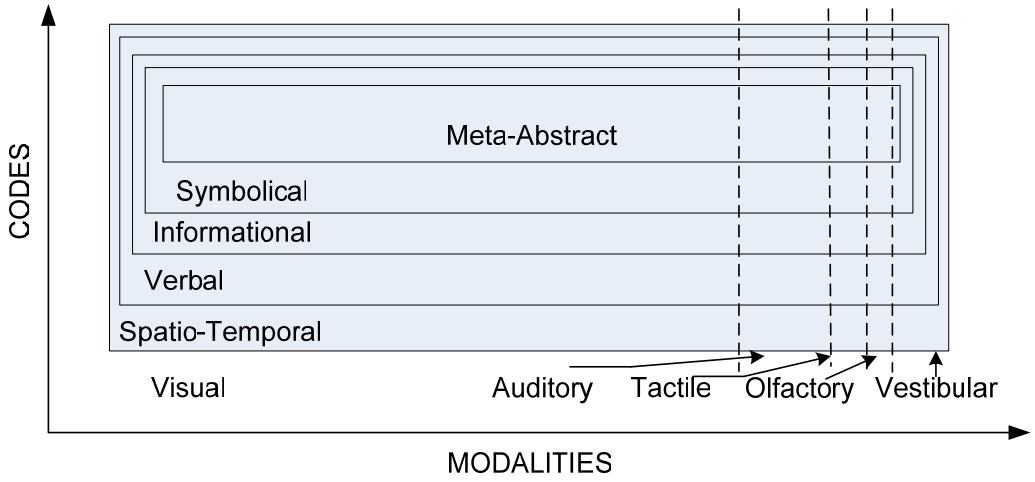


Figure 7: *The elaboration of the now base 'modality' axis combined with the processing codes abstraction hierarchy.*

Processing Sequence

Initially it may appear that the temporal axis that subsumes the processing sequence is somewhat less interesting and less amenable to expansive discourse than the others. However, this is in contradiction to the truth. Indeed, such is the complexity of the way in which the temporal nature of processing occurs that we can only provide a brief commentary here (but see Hancock, Szalma, & Oron-Gilad, 2005). It is necessary to divide this sequence into pre-processing, processing, and post-processing stages. Initially in the pre-processing stage, information is simply the environmental energetic range. This range extends across the electromagnetic continuum and is *a priori* independent of any particular transducer (observer). The energetic representation ascends to *information* within the sensory transduction range of the exposed individual. It is among these *latter* expressions that a stimulus array is available to be selected (and see Gibson, 1979). The principles by which individuals select among a number of possibilities are still being distinguished, however, it is clear that attention is a key element in such selection. Attention is implicated not only in information selection, but

with modulating feedback effects from response. It has been suggested that information flow rate needs to be modulated in the same manner that other physiological systems require stable rates of stimulation. As encapsulated in all the issues about maladaptive levels of workload (e.g., Moray, et al. 1979, Hancock & Meshkati, 1988), it is important for design of human-machine systems to balance information flow (see Hancock & Chignell, 1987). It therefore seems unlikely that optimal real-world design requires visual processing in complete silence or optimal auditory processing to occur in the dark. Indeed, redundancy and cross-talk between channels appears to be an important design issue and, implies crucial permeabilities between Wickens' boxes to generate a full reality. Post processing and feed-forward and feedback loops provide control on the whole sequence and implies the future selection of stimuli and decisions to be made can be largely predicated on past actions since, as we have noted memory is primarily for the future. As this discussion progresses, the box model appears more as a streaming matrix from information to action, rather than repository 'pools' of attention, see Figure 8.

Minsky (1986), opined that the model of realty which connected consciousness was actually emergent property of what he called a 'society of mind.' By this he meant that the unity of consciousness was a functional illusion which only retained a phenomenological wholeness as the by-product of the consensus of multiple contributory processes. In this way, it is not too much to see the present extensions of Wickens' boxes as the psychological expression of these respective mind's 'citizens.' True, one can have the same functional description the interaction of cortical areas neuro-physiological level, but it is our contention that the box model is foundation of its psychological level companion. In the same way that Penfield and Rasmussen(1950) distinguished a sensory homonculus, we believe the elaborated Wickens' model becomes an attentional homonculus (but does not bear the philosophical burden of the naïve interpretation of a homunculus). The sensory homonculus differentiates various parts of the body on the basis of innervation level yet retains the idea of the body as a single,

'whole' entity, so we believe the 'attentional' homunculus performs the same function but in this case the phenomenological whole that is experienced is fundamentally the sense of 'reality.'

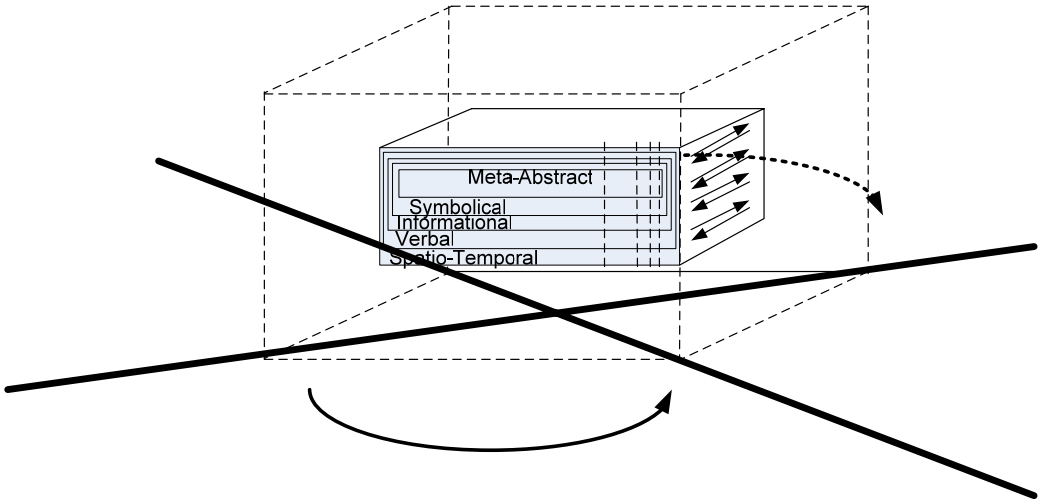


Figure 8: *This representation shows the elaborated 'box model' in the center of the diagram. Surrounding this is an extended box which we take to be the sheath of technology which is being formalized to a greater degree by developments in Neuroergonomics. The large cross on the base, under the boxes, represents a Minkowski space-time diagram. This collapses the three spatial dimensions on one line and the other is the temporal axis. Progress is toward the foreground into a region named the absolute future. Embedded in this future are a series of feed-forward loops, represented as the single, solid arrow. In the quadrant at the back, labeled the absolute past, is a dotted line representative of all feedback loops. Further elaboration of this manner of representing environmental progress can be found in Hancock (1997).*

SUMMARY AND CONCLUSIONS

Despite valiant efforts, the conundrum of attention still remains to be fully resolved. In some ways, the show has now moved on from the original Wickensian notion. Now new largely neuropsychological conceptions have been promulgated and popularized. However, in many ways what seems to represent progress is often simply the same ignorance expressed in another jargon or more colloquially paradigm (this being the fate of most human knowledge of course). Wickens conception was actually one in a long

and tried tradition of psychology (especially the energetic aspects of psychology) in which a unitary concept (e.g., attention, IQ, arousal, fatigue etc.) has been broken into component elements which better explain the nuances of the empirical landscape but often only at the expense of proliferating explanatory degrees of freedom and an eventual fractionation and balkanization that leaves theorists unhappy and experimenters confused and dangling.

Attention is not all of consciousness. The recent pragmatic expression of the operational face of consciousness – situation awareness - has been defined as the external face of consciousness directed beyond the self (see Smith & Hancock, 1995). We know consciousness is more than attention, since evident by stimuli find their way into long-term memory without attention necessarily being directed to them. These characteristics of memory cause surprise when attention is drawn to their subtle penetration (Schacter, 2001). However, much of what the individual recognizes as reality has to be conveyed through the refined portals of attention. And, much as attention can be divided between different elements of the external and internal landscape, we do not experience these divisions as separate phenomenological realities, except in the case of evident mental illness.

That the sensory homunculus describes innervations on various parts of the sensory cortex does not negate the experience of the body as a unified whole. Similarly, that stimulus perception by different sensory modalities in different parts of the sensory environment occurs also does not negate attention or reality as a unified phenomenon. As Gibson (1966) rightly noted, one can erect a theory of perception based upon the exceptions or the illusions, but such a theory would make a very poor heuristic for any organisms' survival or its prosperity in any environment. Similarly, one can believe in diverse attentional pools and the non-conscious processing of stimuli, since the data support such contentions at least to some degree. However, reliance on these would make a poor guide for designing any practical technical system. Perhaps in the end,

Wickens' greatest contribution is that he has served to guide us on this crucial perambulation. In recognizing his unique contribution we now need to proceed more vigorously toward design based upon what humans *should* do, rather than designs based upon economically-driven, efficiency-inspired mandates about what they *can* do. If this transition can be realized then such science will have truly served its ultimate purpose and attention, fractionated or not, will have been directed in the right direction.

Acknowledgements

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, P.A. Hancock, Principal Investigator.

References

Broadbent, D.E. (1958). *Perception and communication*. London: Pergamon Press.

Bronowski, J. (1966). *The common sense of science*. Cambridge: Harvard University Press.

Budiansky, S. (2000). *The truth about dogs*. Penguin: New York.

Damos, D.L. & Lyall, E.A (1986). The effect of varying stimulus and response modes and asymmetric transfer on the dual-task performance of discrete tasks. *Ergonomics*, 29, 519-533.

Freeman, G.L. (1948). *The energetics of human behavior*. Ithaca, New York: Cornell University Press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, Houghton Mifflin.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, Houghton Mifflin.

Hancock, P.A. (1997). *Essays on the future of human-machine systems*. Eden Prairie, MN; Banta.

Hancock, P.A., & Chignell, M.H. (1987). Adaptive control in human-machine systems. In: P.A. Hancock (Ed.). *Human Factors Psychology*. Amsterdam: North-Holland, 305-345.

Hancock, P.A., & Meshkati, N. (1988). (Eds.). *Human mental workload*. Amsterdam: North-Holland.

Hancock, P.A., Szalma, J.L. & Oron-Gilad, T. (2005). Time, emotion and the limits to human information processing. In D. McBride, and D. Schmorrow, (Eds.), *Quantifying Human Information Processing*, in press.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kantowitz, B.H. (1987). Mental workload. In: P.A. Hancock (Ed.). *Human factors psychology*. (pp. 81-121), North Holland: Amsterdam.

Kantowitz, B.H., & Knight, J.L. (1976). On experimenter limited processes. *Psychological Review, 83*, 502-507.

Kinsbourne, M., & Hicks, R.E. (1978). Functional cerebral space: A model for overflow, transfer, and interference effects in human performance: A tutorial review. In: R. Nickerson (Ed.). *Attention and performance VII*. (pp. 345-362). Hillsdale, NJ. Erlbaum.

Minsky, M. (1986). *The society of the mind*. New York: Simon and Schuster.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica, 27*, 84-92.

Moray, N. Johansson, J. Pew, R., Rasmussen, J., Sanders. A.F., & Wickens, C.D. (1979) *Mental Workload, It's theory and measurement*. New York. Plenum Press.

Navon, D. (1984). Resources: A theoretical soupstone? *Psychological Review, 91*, 216-234.

Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. *Psychological Review, 86*, 214-255.

Penfield, W. & Rasmussen, T.L. (1950). *The cerebral cortex of man: a clinical study of localization of function*. New York: MacMillan.

Schacter, D.L.(2001). *The seven sins of memory: How the mind forgets and remembers.* New York: Houghton Mifflin Company.

Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Sivak, M. (1996). The information that drivers use: is it indeed 90% visual? *Perception, 25*, 1081-1089.

Smith, K., & Hancock, P.A. (1995). Situation awareness is adaptive, externally-directed consciousness. *Human Factors, 37*, 137-148.

Stevens, S.S. (1961). To honor Fechner and repeal his law. *Science, 133*, 80-86.

Wachtel, P.L. (1967). Conceptions of broad and narrow attention, *Psychological Bulletin, 68*, 417-429.

Wickens, C.D. (1979). *Introduction to Psychology Seminar*. (Psych 593). University of Illinois, Champaign, IL.

Wickens, C.D. (1980). The structure of attentional resources. In: R.S. Nickerson (Ed.), *Attention and Performance VIII.* (pp. 239-257), Lawrence Erlbaum Associates: Hillsdale, NJ.

Wickens, C.D. (1987). Attention. In: P.A. Hancock (Ed.). *Human factors psychology*. (pp. 29-80), North Holland: Amsterdam.

Wickens, C.D., Mountford, S.J., & Schreiner, W.A. (1981).*Multiple resources, task-hemispheric integrity, and individual differences in time sharing, Human Factors, 23(2), 211-229.*

To Honor Wickens and Restack His Boxes*

P.A. Hancock, J.L. Szalma, and T. Oron-Gilad

MIT² Laboratory
University of Central Florida



Is this a computer I see before me? Early days of Champaign and roses.



Christopher Delos Wickens, Shelagh Joy Mountford and William A. Schreiner, the recipients of the Jerome Ely of the Human Factors Society, at the Seattle Meeting of 1982. The authors award winning 1981 publication was entitled: Multiple resources, task-hemispheric integrity, and individual differences in time sharing, Human Factors, 23(2), 211-229. It was indeed an arduous and enlightening experiment, after all we should know since one of us was a participant.

* While this title may seem somewhat irreverent, its purpose is rather exactly the opposite. Of course, the form is not original, it derives from a famous paper by Stevens in Science: Stevens, S.S. (1961). To honor Fechner and repeal his law. Science, 133, 80-86.

Abstract

Multiple attentional resource theory represents a construct which served to unify two disparate views of human performance capacity. Subsequently, the theory had a particularly strong influence on the practice of human factors, especially in interface development where it arguably remains the strongest behavioral heuristic for design. This confluence of contribution to both fundamental theory and practical application served to facilitate its pervasiveness and impact of multiple resources. Here, we offer elaborations upon this conception with the fundamental purpose being to honor Wickens and seek to build upon what is evidently, his most significant scientific success.

The modal, multiple resource approach is crystallized in Wickens' conception, rendered colloquially as the "box model" (Wickens, 1980). From this conception, which is shown in Figure 1, multiple attentional resources were identified upon the basis of three fundamental dimensions.

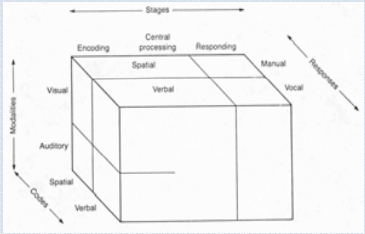


Figure 1: The classic Wickensian 'Box Model' of human attention and design heuristic for information and response distribution for the human operator (after Wickens, 1980).

The prototypical representation of the box model is given in Figure 1. Examples of representative forms in each of the respective cells of the left-facing end axes have also been given by Wickens and are reproduced in Figure 2.

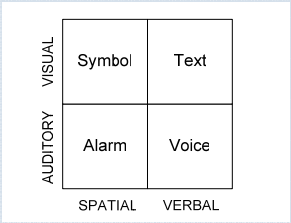


Figure 2: Examples of members in four end cells in the 'Box Model'. A simple auditory alarm tone in the spatial code is expanded into a voice stimulus in the verbal realm. Similarly, a visual symbol is transmitted to text as the change is made from spatial to verbal code.

Let us consider the axis of processing code. In the original formulation, processing code lies on the base axis and is divided between spatial and verbal components. However, this horizontal differentiation implies a qualitative distinction. In fact, one can argue that any verbal representation can be given as an abstraction of the spatial dimension. Let us consider this in a little more detail. In respect of speech (voice), this is a form of verbal code. Similarly, if the verbal code is expressed in the visual dimension, it is viewed as letters, words, sentences, and the like (or more generically – text). However, these are also each only particular forms of geometric ordering in a spatial configuration and therefore again a particular form of abstraction. Thus, if verbal codes are viewed as particular, high-level abstractions of spatial codes (in reality spatio-temporal codes), then we can rotate the boxes so that this is now graphically represented as an hierarchy of such abstraction. This rotation (only for the left-facing 'end' axes) now appear as shown in Figure 3. There is a secondary advantage to this rotation. The new base axis is now modality.

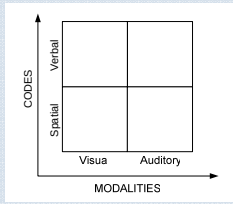


Figure 3: The 'Box Model' rotated through 90 ° in order to emphasize the emergent relationship between the respective codes.

We suggest that if the verbal level is actually an abstraction of the spatio-temporal level then there are other abstractions beyond the verbal itself. We see at least three other levels. They are i) the informational level, ii) the symbolic level and iii) the meta-abstract level. Perhaps an example will be useful in differentiating these levels. Thus, in the ways we have indicated, and of course potentially in ways we have not, the processing code can go well beyond the simple 'spatial' and 'verbal' differentiation of Wickens' original formulation. Whether there is a differentiated visual attentional resource pool for each of these levels, separated say from an auditory attentional resource pool is indeed a debatable issue. In our view this issue (e.g., can one differentiate between seeing and hearing hope?) is one that remains unresolved. We should note that our elaborations are no criticism of Wickens per se, since as a theorist he could only work with the data at hand. Perhaps future advances will provide such information? The addition of these other code extraction levels now provide us with an adjustment to Figure 3 which is now shown in Figure 4. This representation also implies that not all code levels are necessarily of equivalent importance and that there is a necessary nesting involved.

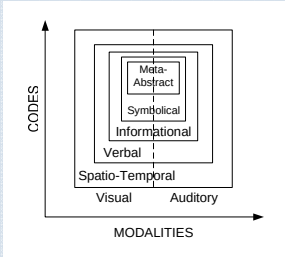


Figure 4: The Processing Codes revised. The codes are now presented as an abstraction hierarchy.

Let us now consider the new base axis - processing modality. In identifying the visual and auditory modality, Wickens specified the two major avenues through which any individual may assimilate sensory information. However, this does not exhaust the number of possibilities. As may be obviously extrapolated, the tactile, olfactory, and kinesthetic senses form a series of modalities and while others such as thermal sensation and pain can also be identified it is important to note that the processing modality axis is not subject to the potential for an infinite regress but rather is limited to the known array of (predominantly human) sensory transduction systems. In the human, vision dominates (Sivak, 1996). Thus the size of the boxes can be *adjusted* to reflect the dominance of and contribution of each particular modality. This is illustrated in Figure 5. Clearly for different organisms, the relative contribution of each modality can change.

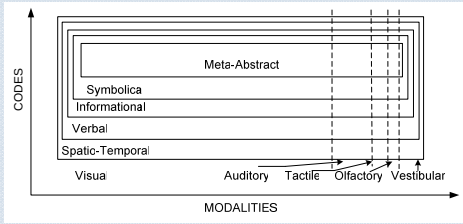


Figure 5: The elaboration of the now base 'modality' axis combined with the processing codes abstraction hierarchy.

The temporal axis subsumes the processing sequence. It is necessary to divide this sequence into pre-processing, processing, and post-processing stages. Initially in the pre-processing stage, information is simply the environmental energetic range. The energetic representation ascends to information within the sensory transduction range of the exposed individual. Attention is implicated not only in information selection, but with modulating feedback effects from response. It has been suggested that information flow rate needs to be modulated in the same manner that other physiological systems require stable rates of stimulation. It therefore seems unlikely that optimal real-world design requires visual processing in complete silence or optimal auditory processing to occur in the dark. Indeed, redundancy and cross-talk between channels appears to be an important design issue and, implies crucial permeabilities between Wickens' boxes to generate a full reality. Post processing and feed-forward and feedback loops provide control on the whole sequence and implies the future selection of stimuli and decisions to be made can be largely predicated on past actions since, as we have noted memory is primarily for the future. As this discussion progresses, the box model appears more as a streaming matrix from information to action, rather than repository 'pools' of attention, see Figure 6.

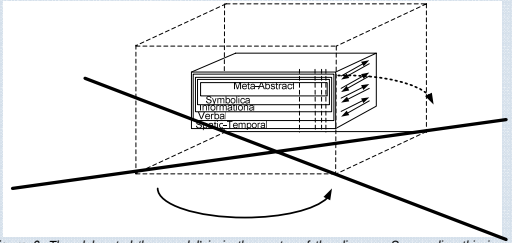


Figure 6: The elaborated 'box model' is in the center of the diagram. Surrounding this is an extended box which we take to be the sheath of technology which is being formalized to a greater degree by developments in Neuroergonomics. The large cross on the base, under the boxes, represents a Minkowski space-time diagram. This collapses the three spatial dimensions on one line and the other is the temporal axis. Progress is toward the foreground into a region named the absolute future. Embedded in this future are a series of feed-forward loops, represented as the single, solid arrow. In the quadrant at the back, labeled the absolute past, is a dotted line representative of all feedback loops. Further elaboration of this manner of representing environmental progress can be found in Hancock (1997).

Summary and Conclusions

The conundrum of attention still remains to be fully resolved. In some ways, the show has now moved on from the original Wickensian notion. Now new largely neuropsychological conceptions have been promulgated and popularized. However, in many ways what seems to represent progress is often simply the same ignorance expressed in another jargon or more colloquially paradigm (this being the fate of most human knowledge of course). Wickens conception was actually one in a long and tried tradition of psychology (especially the energetic aspects of psychology) in which a unitary concept (e.g., attention, IQ, arousal, fatigue etc.) has been broken into component elements which better explain the nuances of the empirical landscape but often only at the expense of proliferating explanatory degrees of freedom and an eventual fractionation and balkanization that leaves theorists unhappy and experimenters confused and dangling.

Attention is not all of consciousness. We know consciousness is more than attention, since evident by stimuli find their way into long-term memory without attention necessarily being directed to them. However, much of what the individual recognizes as reality has to be conveyed through the refined portals of attention. And, much as attention can be divided between different elements of the external and internal landscape, we do not experience these divisions as separate phenomenological realities, except in the case of evident mental illness.

That the sensory homunculus describes innervations on various parts of the sensory cortex does not negate the experience of the body as a unified whole. Similarly, that stimulus perception by different sensory modalities in different parts of the sensory environment occurs also does not negate attention or reality as a unified phenomenon. Perhaps, Wickens' greatest contribution is that he has served to guide us on this crucial perambulation. In recognizing his unique contribution we now need to proceed more vigorously toward design based upon what humans should do, rather than designs based upon economically-driven, efficiency-inspired mandates about what they can do. If this transition can be realized then such science will have truly served its ultimate purpose and attention, fractionated or not, will have been directed in the right direction.

References

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston, Houghton Mifflin.
Hancock, P.A. (1997). Essays on the future of human-machine systems. Eden Prairie, MN; Banta.
Hancock, P.A., & Meshkati, N. (1988). (Eds.). Human mental workload. Amsterdam: North-Holland.
Sivak, M. (1996). The information that drivers use: is it indeed 90% visual? Perception, 25, 1081- 1089.
Stevens, S.S. (1961). To honor Fechner and repeal his law. Science, 133, 80-86.
Wachtel, P.L. (1967). Conceptions of broad and narrow attention, Psychological Bulletin, 68, 417-429.
Wickens, C.D. (1979). Introduction to Psychology Seminar. (Psych 593). University of Illinois, Champaign, IL.
Wickens, C.D. (1980). The structure of attentional resources. In: R.S. Nickerson (Ed.), Attention and Performance VIII. (pp. 239-257), Lawrence Erlbaum Associates: Hillsdale, NJ.
Wickens, C.D. (1987). Attention. In: P.A. Hancock (Ed.). Human factors psychology. (pp. 29-80), North Holland: Amsterdam.
Wickens, C.D., Mountford, S.J., & Schreiner, W.A. (1981).Multiple resources, task-hemispheric integrity, and individual differences in time sharing, Human Factors, 23(2), 211-229.

Acknowledgement

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, P.A. Hancock, Principal Investigator. The work was also facilitated by a number of additional Grants and Contracts and their support of the Laboratory is indeed, gratefully acknowledged.

The Role of Hedonomics in the Future of Industry, Service, and Product Design: Panel Overview

Tal Oron-Gilad, Chair and P.A. Hancock, Co-Chair

University of Central Florida
Orlando, FL

Hedonomics has been defined as "*the branch of science which facilitates the pleasant or enjoyable aspects of human-technology interaction*." Hedonomics in general is a fairly new area in research and it is evidently new in ergonomics, human factors, usability and human-computer interaction (HCI). During the last ten years there has been a rapid growth in research about affect and pleasure. Considering the lack of interest from the psychological community during much of the 1900's this comes as a surprise. Affective evaluations provide a new and different perspective in Human Factors Engineering. It is not how to evaluate users – it is how the user evaluates. The research on hedonic values and seductive interfaces is in fact a welcome contrast to issues of safety and productivity, which have dominated human factors and ergonomics (HF/E). In order for industry to adopt these user-evaluating principles (as opposed to user-evaluation principles) it has to be convinced of their practical effectiveness in the workplace. This panel will focus on this issue and the role of Hedonomics in future work environments and products.

INTRODUCTION

Hedonomics has been defined as "*that branch of science which facilitates the pleasant or enjoyable aspects of human-technology interaction*" (Hancock, Pepe, & Murphy, 2005). Hedonomics is a fairly new area in research and it is clearly new in ergonomics and HCI. During the last ten years there has been a rapid growth in research concerning affect and pleasure. Considering the lack of interest from the psychological community during much of the 1900's this comes as something of a surprise. Affective appreciation is of course not new – just the formalization of the relevant research. People have affective reactions towards tasks, artifacts, and interfaces all the time. These are caused by design features which operate either through the perceptual system (looking at) or from a sense of controlling (touching and activating) or from reflection and experience.

Affective evaluations provide a new and different perspective in Human Factors Engineering. It is not how to evaluate users – it is how the user evaluates. The research on hedonic values and seductive interfaces is in fact a welcome contrast to safety and productivity, which have dominated human factors and ergonomics. Consequently, emotions and affect have received increasing attention over the last recent years. Approaches to emotions and affect have been studied at many different levels, and several models have been proposed for a variety of domains and environments.

In order for the industry to adopt these user evaluating principles (as opposed to user evaluation principles) it has to be convinced of their practical effectiveness in the industrial commercialized realm. Dr. Oron-Gilad will provide this top-down perspective on the evolution of ergonomics and the role of Hedonomics in the future of industry and service, in a broader way than just in the design of products for leisure and fun. Dr. Szalma will provide the psychological background in motivation and emotion relevant for the applications of Hedonomics in the human factors context. Dr. Helander will talk about theory and modeling. Hedonomics calls for innovative theory and novel affective evaluations which will provide a new and different perspective in Human Factors Engineering. It is not how to evaluate users – it is how the user evaluates, a very different approach than the approach human factors have addressed so far. Dr. Khalid will talk about methodology development for the evaluation of affective and pleasurable design. Finally, Dr. Meyer will discuss some directions in which existing methods for task analysis need to be changed to accommodate a Hedonomic view of system design. Dr. Hancock will discuss the general principles of Hedonomics. Specifically, he will focus on the fundamental goals which seek pleasurable interaction but not achievement without effort or the simple alleviation of all need. Dr. Hancock will discuss these issues with respect to major, influential philosophical positions expressed by More, Hobbes, Bentham, and Teilhard de Chardin, among others. The proposition will be considered that Hedonomics is actually a facet of technology that can solve contemporary political conundra concerning the asymptote of global political organizations expressed as converging structures across the nations of the developed world (Fukuyama, 1992).

The discussion will be directed toward several open questions raised by the panelists and the chair: Can Hedonomics be applied in the work environment and how can we 'sell' Hedonomics as part of the evolving necessity of increasing productivity in workplaces? In

addition, theoretical models and approaches toward measurement and analysis of human reactions to affective and pleasurable design will be discussed. The discussion will also attempt to encounter ways to alter traditional design methodologies.

SUMMARIES FROM PANELISTS

A Historic Perspective on the Evolution of Ergonomics

Tal Oron-Gilad

Institute for Simulation and Training
University of Central Florida

From an historic perspective, safety, ergonomics, and usability were always adopted by industries, workplaces, and manufacturers after they have proven to increase productivity and profit. Furthermore, they each build upon each other and represent a hierarchy of needs in the workplace, i.e., safety is the most fundamental need and usability has been the latest "top level" of the pyramid. Unfortunately, operators' well being was and still remains rarely a matter of concern per se. Human Factors has traditionally followed this technology and performance-driven approach by facilitating better designs for systems and workplaces to achieve higher productivity goals. Similarly, Hedonomics will not be adopted by industries and workplaces unless it proves to follow this productivity increase trend and become the next novel-innovative top hierarchy of the pyramid. Can Hedonomics be applied in the work environment and how can we 'sell' Hedonomics as part of this evolving necessity?

Application of Theories of Motivation and Emotion to Hedonomics

James L. Szalma
University of Central Florida

Theories of perception and cognition are routinely employed in ergonomics and human factors. In a similar fashion, hedonomics, as it develops, will likely draw on theories of motivation and emotion for advancing the goal of facilitating the development of convivial (Illich, 1973) and pleasurable technologies. Although there is an overabundance of theories of motivation and emotion, general themes emerge that are relevant for a mature hedonomic science. Thus, most theories incorporate goals (e.g., Locke & Latham, 1990), often organized into hierarchies, as a central driver of motivation and emotional states. Relevant issues include how goals are selected, the consequences for meeting or failing to meet those goals, and the rate of progress toward goal states. Another central theme is that of appraisal, in which individual's motivational and emotional states are determined by their subjective evaluation of events (Lazarus, 1991; Scherer, 1999). These appraisals are influenced by estimates individuals make regarding the relative likelihood of outcomes (expectancies) and the personal relevance of those outcomes (valences and values; e.g., George & Jones, 1997; Vroom, 1964). Another theme is that of self-regulation of effort and goal pursuit, which is concerned wit the mechanisms by which individuals regulate their emotional states (e.g., Carver & Scheier, 1998).

For motivation in particular, a final theme is that of psychological 'needs,' that drive behavior. A particularly useful theoretical perspective that emphasizes needs is self-determination theory (Ryan & Deci, 2001; Deci & Ryan, 1985; 2000). From this perspective there are three organismic needs that are essential for facilitating intrinsic motivation for task activity and the positive affect that can accompany such states. These needs are for competence (self-efficacy; see also Bandura, 1997), autonomy (personal agency, not independence *per se*), and relatedness. For hedonomic interventions to succeed, they must facilitate to the greatest degree possible genuine fulfillment of these needs. From this perspective, a simple usability approach is likely to fail, since it is based upon user preferences rather than whether the technology facilitates the individuals needs for competence, autonomy, and relatedness. Further, for hedonomic design to reach its fullest potential in work environments, the organizational/social context must be considered part of the system to be designed. This will require a comprehensive socio-technical approach the design process.

In regard to emotion, an additional central theme of research has been the nature and taxonomic structure of the 'basic' emotions (e.g., see Ekman & Davidson, 1994). Multiple theoretical models exist, but a potentially useful framework for hedonomic design is the two dimensional affective space described by Watson and Tellegen (1985). In their model, affective states are organized along two orthogonal dimensions, positive and negative affect. Thus, individual states can be high on both the positive and negative affect dimensions (i.e., highly activated and engaged with one's environment) or low on those dimensions (i.e., inactive and disengaged from the environment). Watson and Tellegen's (1985) representation can serve as an initial guide by identifying those affective states clustered around the high positive/low negative portion of affective space, although in some applications other sets of emotions (e.g., high positive/high negative) might be more desirable.

As in other areas of human factors, purpose will be the central issue in the application of motivation and emotion theory to hedonomic design (Hancock, 1997). How one applies these theoretical models (and which theories are adopted) and which hedonomic goals are selected will depend on the explicit or implicit purpose for the design. A key issue to be resolved in future considerations is whether the general purpose for

hedonomics should be the promotion of pleasant and enjoyable human-machine interaction, or the goal of promoting well-being and using technology to facilitate well-being. This, in turn, depends in part on whether one's approach to the study of well-being is based on pleasure (hedonism) or self-actualization (eudaimonism; see Ryan & Deci, 2001). While the two goals will in many cases be concordant, there will also be many instances in which they will be contraindicated (i.e., increase the alienation and of the individual and exchanging short term pleasure for long term well-being). In such cases, which position prevails will ultimately depend on the intention of those who create and operate the technology.

E**valuation of Affective Design. A Problem of Measurement.**

Halimahtun M. Khalid

Damai Sciences Sdn Bhd, Kuala Lumpur, Malaysia

Traditional cognitive approaches to product usability tend to underestimate the role of user emotions and customer needs. Not surprisingly, the success of a product in the marketplace may be determined by its aesthetic appeal, the pleasure it creates and the satisfaction it brings to the user. Such user experience encompasses emotions and expectations in relation to the context of use of the product. Emotions influence how a user interacts with the product. In the interaction, emotions accompany thinking.

A framework for analysis of user emotions in relation to the designer's environment is described. The framework is part of the product development life cycle, where user needs and measured emotions are fed early in the design process. The goal is to achieve a pleasurable and satisfying product.

A major research problem in the science of Hedonomics is to develop methods for measuring emotions. There are five criteria to consider in measuring and evaluating emotions, namely, dynamics, context, reliability, validity and measurement error. The literature has documented various methods derived from Psychology and Marketing. Select subjective and objective methods that are commonly used in evaluating user emotion to an artifact are discussed, with illustrations from various product applications.

Measurement of emotions may be difficult due to the indefinite relationship between an emotion and its behavioral expression. This is because the mechanics of hedonics are not fully understood. Further research is needed to develop expressions of emotions that are quantifiable, so that they can be verified easily. Examples from computing and entertainment are highlighted.

Separating emotion from cognitive functions does not seem helpful from a research perspective or from a design perspective. Instead, an integrated view of emotion and cognition is taking hold; not only in neuroscience but also in product design. Additionally, a product should be designed to support customer needs, including the customer's persona or personality attributes. This can be done by providing flow – or ease of use – and inducing feelings – or emotions – in interacting with the product

In the future, when we understand the "physics" of emotions and what makes people happy, there will be significant rewards in terms of monetary benefits for developers of pleasurable products, as well as many "feel good" customers.

Task Analysis for Fun

Joachim Meyer

Department of Industrial Engineering
Ben Gurion University of the Negev
Beer Sheva, Israel

Human Factors has traditionally adopted a teleological stance that views activities as goal-driven and aimed to perform a task. Systems accordingly need to be designed to facilitate the achievement of these goals. However, people quite often engage in an activity and use an inefficient system because they enjoy the use itself. For instance, drivers may sometimes enjoy a difficult driving task, such as driving along a winding rural road in a car with manual transmission. In such cases, system designers need to focus on enhancing user enjoyment of the activity. Standard human factors methods, including conventional tools for task analysis, are ill suited to provide appropriate design guidelines when user enjoyment is the principle purpose of using the system. An alternative analytical process needs to be developed that relies on the analysis of emotional, sensory and perhaps sensual experiences during system use, in addition to traditional aspects of task analysis.

Hedonomics: Beyond *Dulce Far Niente*

P.A. Hancock

Department of Psychology, and
Institute for Simulation and Training
University of Central Florida
Orlando, FL

Some major religions tell us that all of life is suffering and that the alleviation of that suffering is to be had only in the renouncement of all desiring. Other theologies offer us glimpses of paradise, promising either the full surcease of sorrow (nepenthe) or residence in place of eternal bliss. These respective visions are not

without their critics (Twain, 1962). Modern technology promises to create for us some of these visions of the afterlife here, now, on earth. But what of these paradises, of what are they composed? Many see these visions in terms of an end to need. In modern psychological terms, this may be expressed as the satiation of all levels of Maslow's hierarchy. Others see this vision in terms of the end of effort. In essence, this represents the vision of "*dulce for niente*" – life without care. In contrast, we do not see the future of Hedonomics as seeking to achieve either the total relief of human need or *dulce for nietne*. Rather, we see Hedonomics as a branch of the science of work that aspires toward enjoyable but challenging and explorative interaction with technology. We therefore see pleasure as a process rather than a state. Thus the fundamental dynamic of that process - time - is a critical and central facet of Hedonomic science. Immediate questions involving rhythms of engagement, waves of attention, and intrinsic and extrinsic activity cycles now come immediately to the fore. We claim that pleasurable interaction with technology itself possesses these intrinsic 'waves' of synergy and thus can never be a simple static condition. One critical companion co-variant of Hedonomics is the allied and linked concept of 'individuation' (Hancock, 2003) In human interaction with technology, since technology's dominance evident in the industrial revolution, we have seen an evolutionary progression from the straight-jacket of mass production characterized by the 'one size fits all' Henry Ford type conception, through a phase of technological adaptation in which systems permit users to make a limited range of personal adjustments based upon their own preference. The next, and evident phase in this line of progress is one of technology's personal adaptation or 'individuation' in which the individual basically crafts their own work and environment, supported by flexible, 'convivial' technology (Illich, 1973) The prime, common function between individuation and Hedonomics is personal autonomy (control). The next step for these companion conceptions is to understand the nature of challenge and the regulation of motivation and mental load in relation to the obligatory, task demands of any endeavor. The first step is a careful exploration of the degree of necessary linkage between the physical task domain and the expression of that domain in the human-machine interface. As a polemic assertion, we claim there is no necessary connection. This being so, the nature of computer-mediated work is ultimately malleable and therefore Hedonomically manipulabile. Although there is evidence for convergence of evolution driven by the interface configuration (Hancock, 1996), Hedonomic principles can counter this convergent trend. In such a battle, we have to understand the commonalities of what things, events, structures, and processes make many people happy? (and see Bentham, 1822)

REFERENCES

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman and Company.

Carver, C.S., & Scheier, M.F. (1998). *On the self-regulation of behavior*. Cambridge: Cambridge University Press.

Deci, E.L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

Ekman, P., & Davidson, R.J. (1994, Eds.). *The nature of emotion: Fundamental questions.* Oxford: Oxford University Press.

George, J.M., & Jones, G.R. (1997). Experiencing work: Values, attitudes, and moods. *Human Relations*, 50, 393-416.

Hancock, P.A. (1997). *Essays on the future of human-machine systems.* Eden Prairie, MN: BANTA.

Illich, I. (1973). *Tools for conviviality*. New York: Harper & Row.

Lazarus, R.S. (1991). *Emotion and adaptation.* Oxford: Oxford University Press.

Locke, E.A., & Latham, G.P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice Hall.

Ryan, R. M., & Deci, E.L. (2000). Self-determination theory and the facilitation fo intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78.

Ryan, R.M, & Deci, E.L. (2001). On happiness and human potential: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, 52, 141-166.

Scherer, K.R. (1999). Appraisal theory. In: T. Dalgleish and M. Power (Eds.), *Handbook of cognition and emotion.* Chichester: Wiley.

Vroom, V.H. (1964). Work and motivation. New York: Wiley.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219-235.

Exploiting the human in human machine systems



Tal Oron-Gilad, Ph.D
Peter A. Hancock, Ph.D.
James L. Szalma, Ph.D.

MIT² Laboratory

Institute for Simulation & Training
and Department of Psychology
University of Central Florida, Orlando FL

Stress in the Computing Environment
1st Annual Augmented Cognition International Conference
Las Vegas, Nevada, July 22-27 2005

What is Augcog?

Augcog finds ways in which technology can serve to augment the inherent capacities that people already possess

- human centered (e.g., Billings, Woods)
- individuation (Hancock, 2003)

User centered approaches are more difficult to implement than technology-driven designs.

Are they in practice and in principle also inherently slower or less efficient?

Billings, C.E. (1996), *Aviation Automation: The Search for a Human-Centered Approach*, (Mahwah, NJ: LEA).
Hancock, P.A. (2003). Individuation: Not merely human-centered but person-specific design. *Proceedings of the HFES*, **47**, 1085-1086.

Human-Machine Symbiosis

It is becoming evident in the user-centered form of symbiosis that:

a) the designer is grappling to understand the nuances of the user (emotion, individuation)

b) the user is the limiting function (controlling/dominant factor)

"Hugh Patrick Ruffell Smith, admitted in 1949 that, "Man is not as good as a black box for certain specific things. However, he is more flexible and reliable. He is easily maintained and can be manufactured by relatively unskilled labour." This maxim is as true today as when he formulated it almost 50 years ago" (from Billings, 1997)

Can we improve the situation?

1. Automate more - automate more wisely
   Timing in automation is under explored
   Individual characteristics are not sufficiently implemented

2. Explore what it is about human beings that continues making them relatively slow?

What are the limits of human information processing capacity

- 7±2 (Miller, 1956)
- ATC 17 aircraft per section
- 4-way interaction
  (Halford et al., 2005)

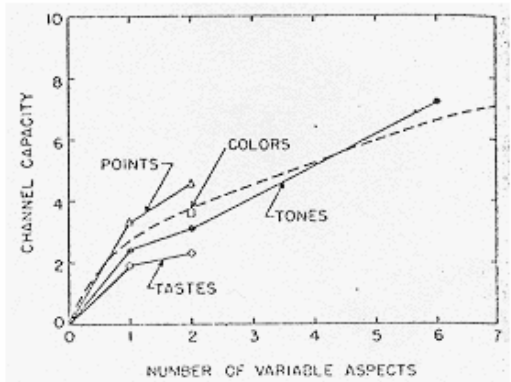


FIG. 6. The general form of the relation between channel capacity and the number of independently variable attributes of the stimuli.

1. Are these **inherent barriers** or are these **intrinsic barriers**?
2. Are they simply cases that relatively few people can solve?
3. What is the role of **training** to solve these barriers?
4. What is the role of **design** to solve these barriers?

G.A. Miller (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information, *Psychological Review, 63*, 81-97. Available online at http://psychclassics.yorku.ca/Miller/

Inherent Barriers or Intrinsic Barriers



Three Spanish parachutists crossed the strait of Gibraltar in freefall, breaking two world records. They jumped over Morocco from the extreme altitude of 13,000 meters (43,000 feet), out of a military aircraft.

With an average forward speed of 208 kph they were freefalling for roughly 6 minutes, and flew the approximately 20 kilometers between Morocco and the Spanish mainland.

Only after completely crossing the strait of Gibraltar did they open their parachutes and land on a beach. "The worst was the cold", the men reported after landing. At 12,000 meters and higher, icy -50 degrees celsius temperatures prevail. They also noted that **the six minutes in freefall seemed like an eternity.**

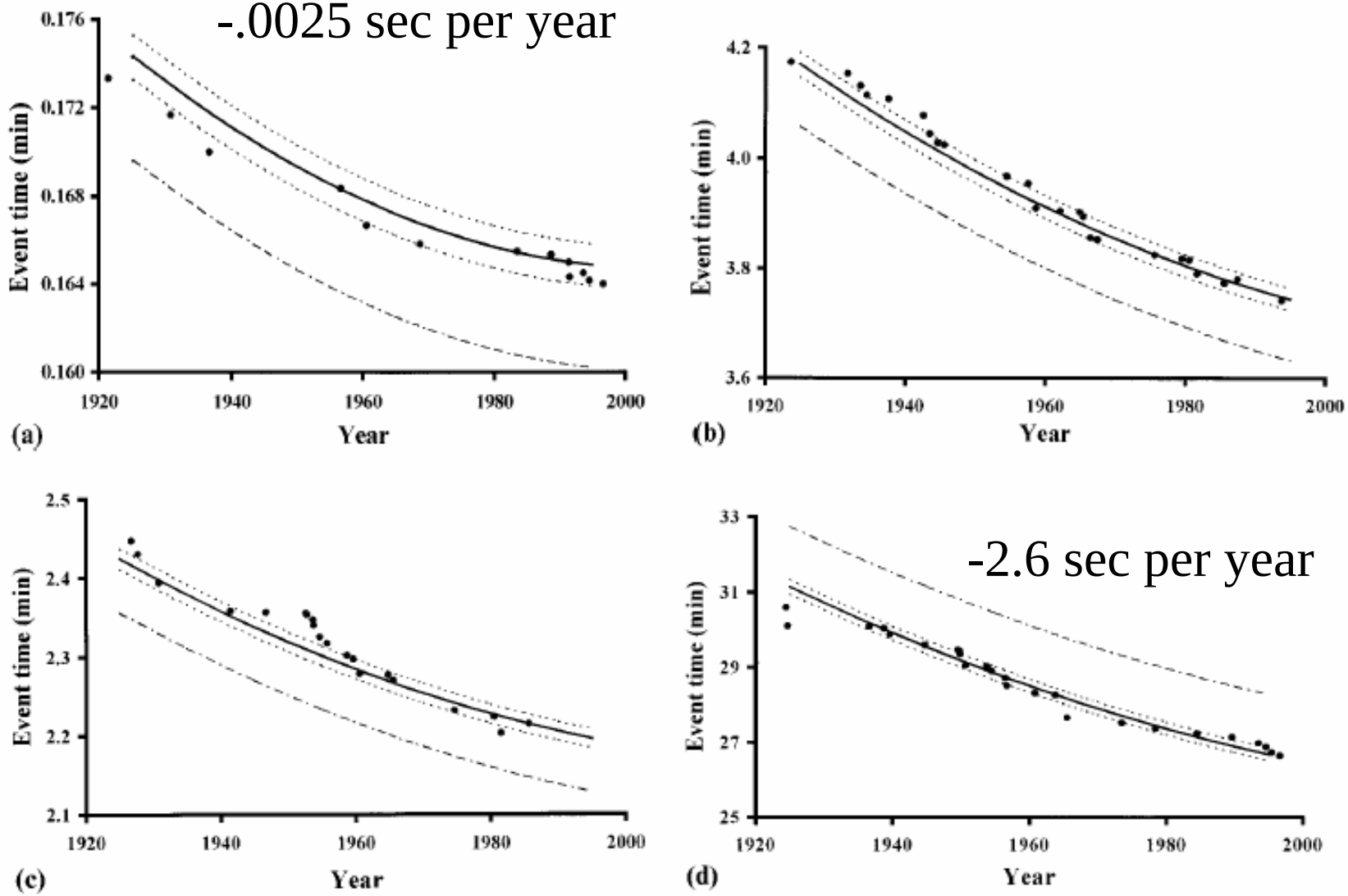


Figure 7 Event time versus year for: (a) the men's 100-m run; (b) the men's one-mile run; (c) the men's 1-km run; (d) the men's 10-km run.

Katz & Katz, 1999

Super athletes run as if they have identical bodies
Aerobic + anaerobic energy utilization in running remained constant over the years

Cases that relatively few people can solve

Autistic Savant

The human brain can perform extremely fast and complicated calculations that still well exceed computer capacities.

Yet,

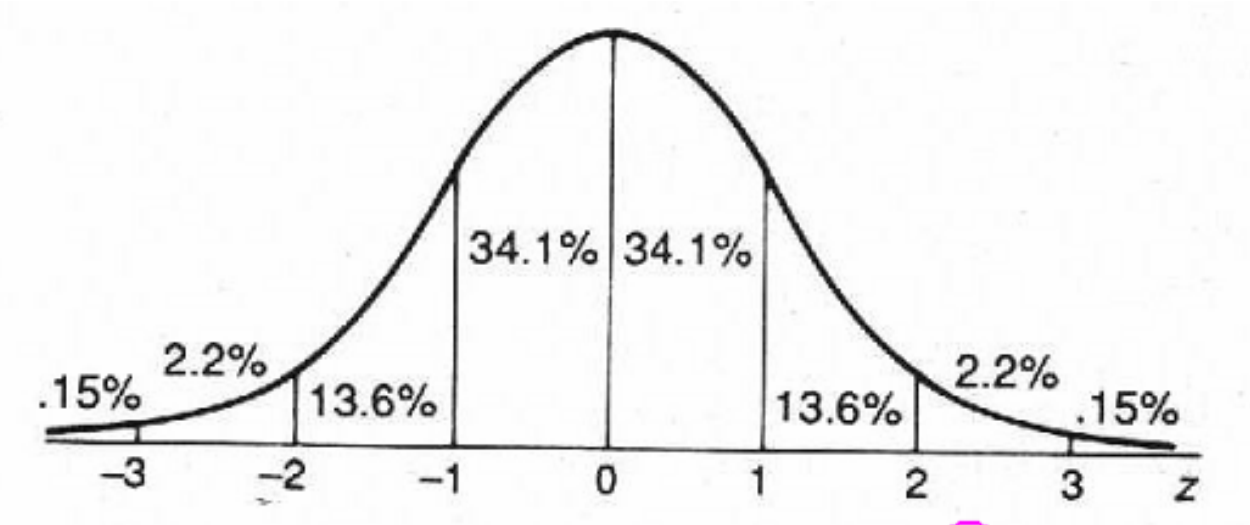
to obtain those capabilities you seem to have to sacrifice other capabilities (usually social capabilities).

However,

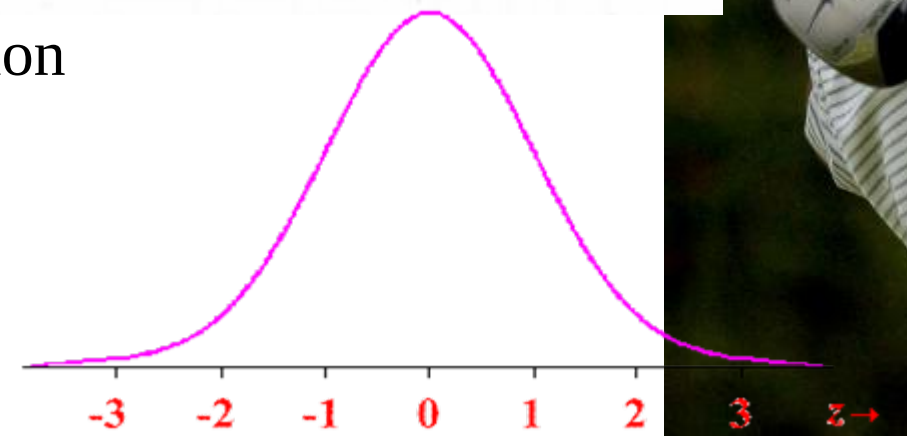
there is no argument that the brain is capable of much greater range of operations and potentially a much faster way of processing.

Autistic Savant refers to individuals with autism who have extraordinary skills not exhibited by most persons. The estimated prevalence of savant abilities in autism is 10% (1% among non-autistic population). Edelson, 2001

Cases that relatively few people can solve



Population distribution

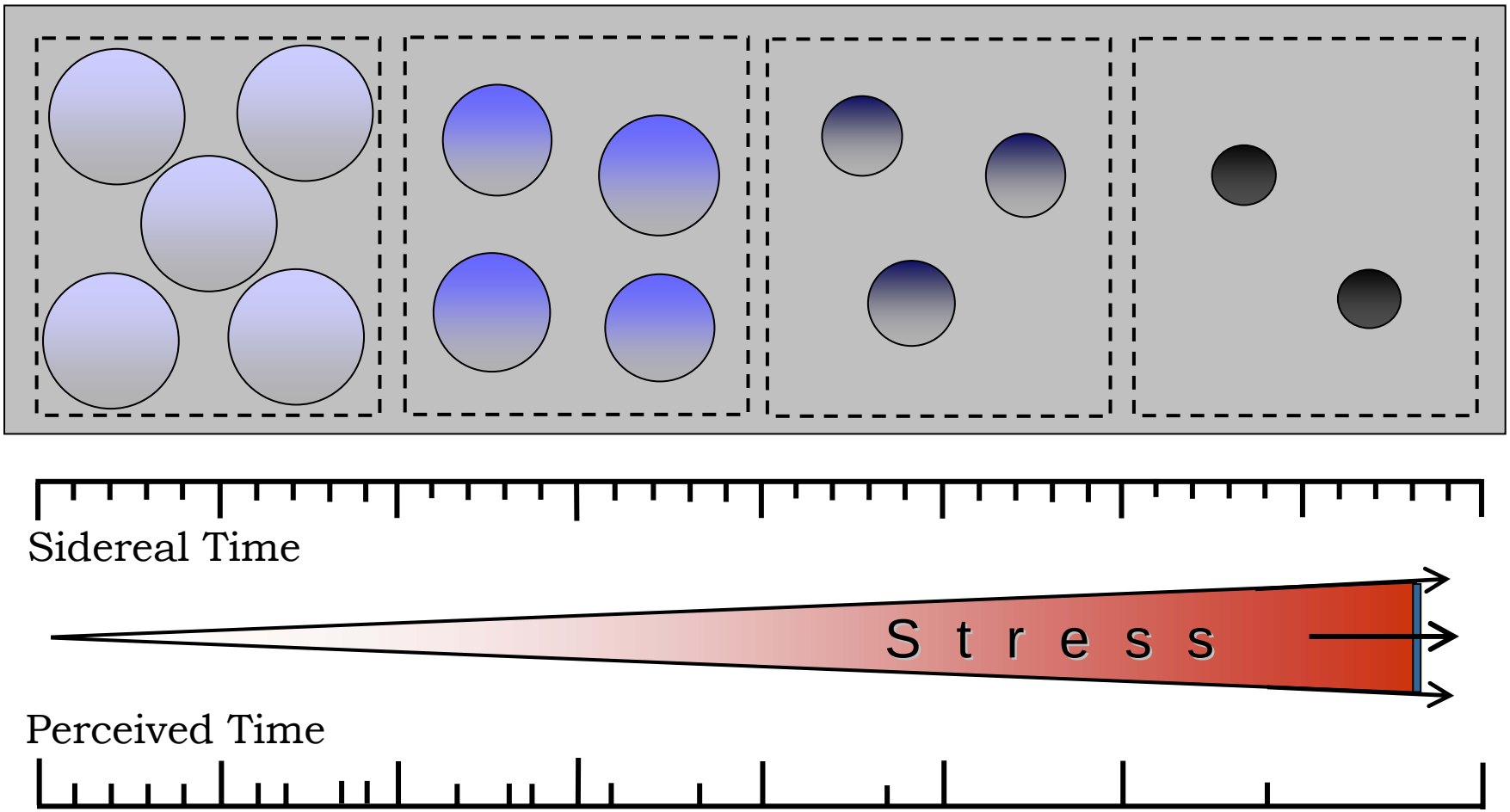


Individual's distribution



6σ=0.0003 3.4 per million
7σ 38 per billion

Performance Under Stress



Hancock, P.A., & Weaver, J.L. (2005). Temporal distortions under extreme stress. *Theoretical Issues in Ergonomic Science*, 6(2), 193-211.

Performance under stress: Sometimes Time is Not Just Time

"During a violent shoot out, I looked over, drawn to the sudden mayhem, and was puzzled to see beer cans slowly floating through the air past my face. What was even more puzzling was that they had the word *Federal* printed On the bottom. They turned out to be the shell casings ejected by the officer who was firing next to me."

"I was talking when I heard a screech and a bang. Then we slid off the road and the ground was flying toward us. It was all a blur and going round and round when it came to a halt the roof was right near the floor. I managed to crawl out of a gap where the windows should have been. People had been thrown out through the windows. It was horrible. I tried to help where I could but there wasn't much we could do. There were people trapped but we couldn't pull them out. It is a difficult thing to describe. The scene was not nice. People have died and are very badly hurt. No-one wants to see that. It happened hours ago but it still only seems like ten minutes."

Barriers that Training can solve

Chase and Ericsson, 1982
(7 to 79 digits, 230 hours of practice over 20 months)

Skilled Memory Theory, three basic principles:

1) information is encoded into LTM in terms of prior knowledge
2) a separate LTM knowledge structure (a meta-structure) is used to keep track of the order in which the information was presented
3) encoding and retrieval speedup with practice.

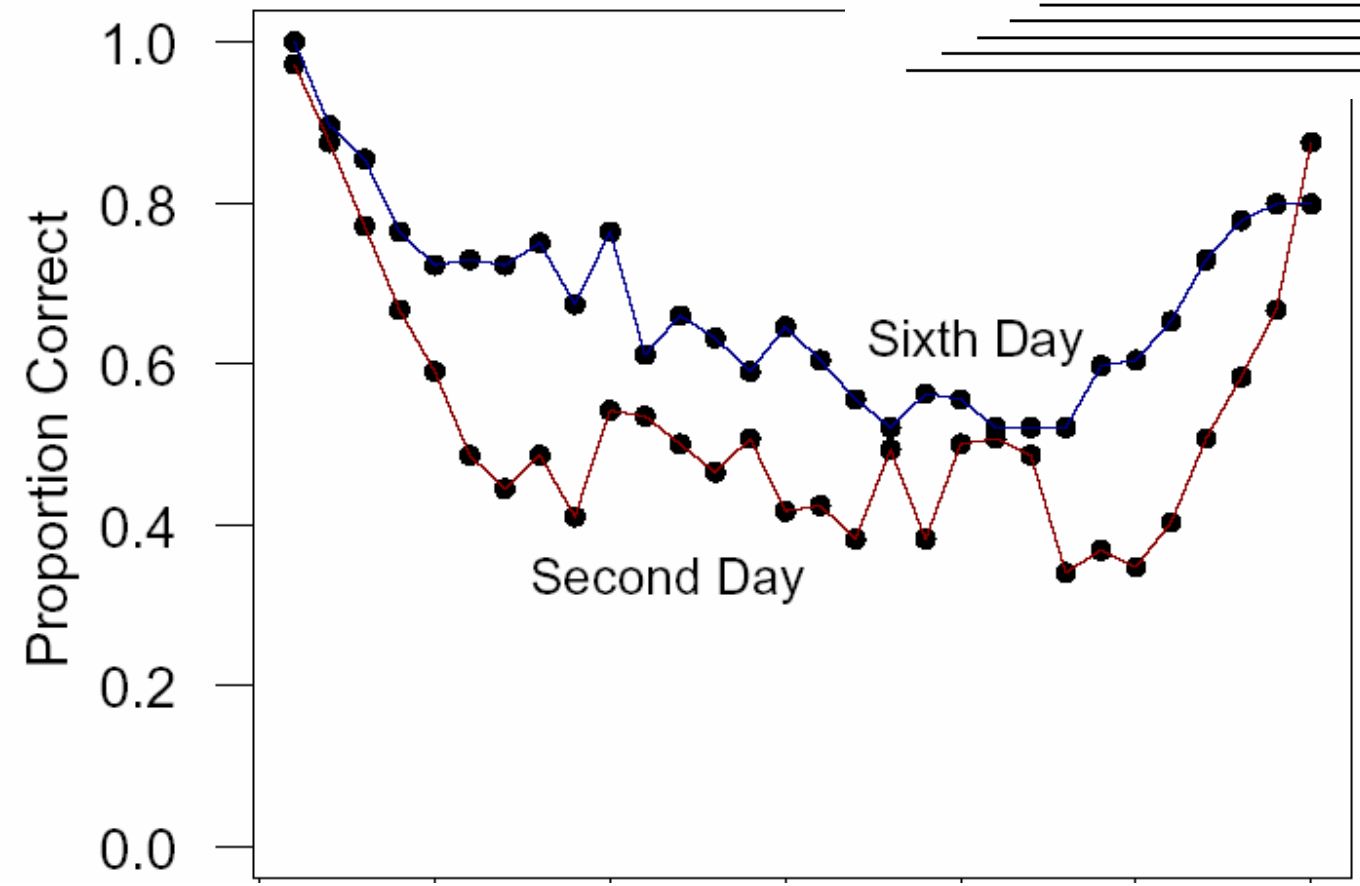
Ericsson and Staszewski (1989)

LT-WM

- The ability to **anticipate** future retrieval demands
- Cue-based retrieval without additional encoding

Ericsson, K. A. & Kintsch, W. (1995)

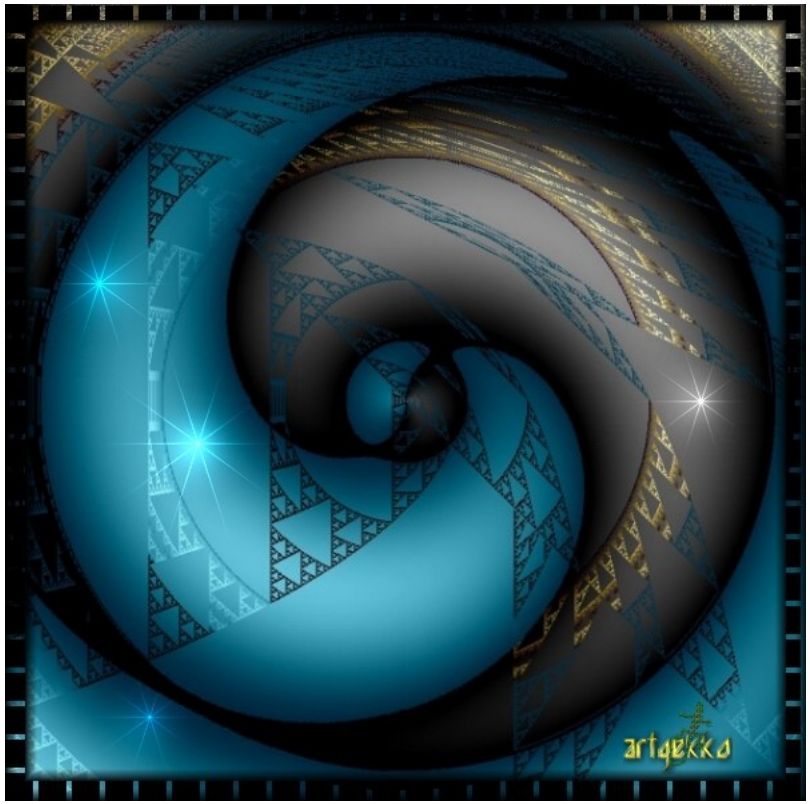
Barriers that Training can solve



Rouder, Moray, Cowan & Pfaltz, Learning in unidimensional absolute identification

Barriers that Design can solve

Information is not Linear in Bits



How Long Should a Man's Legs Be? -Just Long Enough to Reach the Floor.

How many "bits" In a Piece of Information? – Just Enough to Satisfy the Contextual Constraint

Time & Time Perception

Are human beings limited in time as they are limited in space (7 ± 2 in time????)

If so then

When we fracture the constraints we can probably do the same thing in time that Chase & Ericsson, 1982 among others have done in memory

In sum…

- The intrinsic problems can be set by understanding how people perceive time
- The spatial domain has been well minded, the temporal domain remains unexplored.
- We have several ways to progress….

Acknowledgments

This work was prepared under Grant #NBCH 1030012 from the Defense Advanced Research Projects Agency, (DARPA) via the Department of the Interior. The Grant Monitors were Cmdr Dylan Schmorrow and Cmdr Russ Shilling, P.A. Hancock, Principal Investigator. We would like to acknowledge the very fruitful interchange of information and opinions with Dr. Rene de Pontbriand in the formulation of the present ideas. The report itself was under the direction of Dr Dennis McBride of the Potomac Institute.

This work was facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, Dr. Elmar Schmeisser, Technical Monitor, P.A. Hancock, Principal Investigator

The views expressed in this work are those of the authors and do not necessarily reflect official Army or Department of Defense policy.

ARE WE READY TO CONSIDER INDIVIDUAL DIFFERENCES IN HUMAN CAPABILITIES IN OUR WORKPLACE DESIGNS?

Panel Organizer and Chair
Krystyna Gielo-Perczak
Liberty Mutual Research Institute for Safety
Hopkinton, MA

Panel Co-Chair
Waldemar Karwowski
Center for Industrial Ergonomics
University of Louisville
Louisville, KY

Panelists

Peter A. Hancock
Department of Psychology
Director, MURI-OPUS Laboratory
University of Central Florida
Orlando, FL

Waldemar Karwowski
Center for Industrial Ergonomics
University of Louisville
Louisville, KY

Krystyna Gielo-Perczak
Liberty Mutual Research Institute for Safety
Hopkinton, MA

William S. Marras
Institute for Ergonomics
Department of Industrial and Systems Engineering
The Ohio State University
Columbus, OH

Raja Parasuraman
Department of Psychology
George Mason University
Fairfax, VA

INTRODUCTION

At the beginning of the 21st century, individual differences in performance are still not being included in workplace design. Most often they are ignored, minimized or not considered at all. Since the nature of individuality has so many aspects, not considering its complexity is a constant routine for many human factors practitioners and researchers. Still, we record a high number of musculoskeletal disorders and have the desire to lower these statistical records. Advanced analysis of workplace design points to the need for a broad approach to individual differences in performance, taking into account their many aspects, including physical and mental capabilities. The goal of this panel is to generate an environment that promotes the exchange of ideas between neuroergonomics, biomechanics, medical imaging, genetics, physiology, psychology, sociology, neuro-ecology regarding ergonomic design as an integration of the individual differences within human capabilities.

As the title of this session indicates, we are currently in the process of recording individual differences in human capabilities. However, are we ready to consider these individual features in workplace design? How far are we going with the understanding of individual human performance? What aspects of this analysis are at the level of laboratory data collection or modeling, as a extension of a classical cognitive engineering framework? To address these questions, this panel has been assembled to discuss issues with regard to what individual difference variables are most accountable to investigation within human factors.

The discussion begins with an evolutionary statement of why recognizing individual differences in human capabilities and limitations is necessary. A transitional thought of "designing for the mean rather than the individual" will be elaborated on in the next step of the discussion, pointing to a biomechanical approach respecting geometric individual differences. Then, the discussion turns to presenting the recent breakthrough results of neuroergonomic studies of cognitive abilities. The "gene dose" of certain polymorphisms (variants of genes) that have been identified in the human genome can be shown to predict individual differences in working memory capacity. The next area being investigated demonstrates the key role of personality factors in work environment interactions. The classical cognitive engineering framework is extended through the neuro-ecological approach, including personal human attributes important in exercising control over the work environment. And, finally, the last panelist draws attention to the need for combining all the aspects of analysis of individual differences in performance. The goal of this panel session will be to attempt to bridge the diverse remarks made by the panelists and identify a common theme for investigating individual differences in human factors research.

PANELIST'S ABSTRACTS

Peter A. Hancock
University of Central Florida

Our Natural Evolution Toward Individuation

At one time, individuals were very much forced to fabricate their own tools for themselves. At this early stage of development all such fabrication was crude but certainly centered on each individual user. The growth of tool sophistication was almost inevitably accompanied by specialization and a concomitant dissociation between manufacturer and users. Rapidly, manufacturers understood the economic advantage of designing for the mean rather than the individual which only persisted in the cases of exceptionally powerful individuals. The start of the twentieth century saw the height of this trend. Now we have move to the stage of development in which systems, products, etc can adapt to the user. The next natural step in evolution is not adaptation but a return to individual customization, or 'individuation' as I have termed it. My contribution to the present panel will be an examination of this historical and evolutionary trend and its implications for future Human Factors.

Krystyna Gielo-Perczak
Liberty Mutual Research Institute for Safety

Individual Differences in Strength Performance Restrained by Joint Geometries

Occupational ergonomics and safety considerations relating to the human body require a new biomechanical approach that combines the geometric complex nature of the human joints with the requirements of the workplace in order to explain the mechanism of musculoskeletal disorders. The commonly used standard recommendations in workplace design for maximum acceptable load are still too high and should be revised downward. Before the standards can be changed we should take into account individual body strength by examining effectual factors on individual strength performance.

Individual body strength is limited by the capability of their joints. One of the recent studies related to upper extremities presents the influence of glenoid geometry on shoulder static strength during one-handed pushing. The study revealed the correlation between the surface area (called the area of glenoid asymmetry as represented by a product of glenoid height and the difference between the lower and upper depths of the glenoid), and shoulder joint strength during one-handed pushing. The defined asymmetry of the glenoid fossa might have influence on the individual maximum load of the glenohumeral joint that can be generated during pushing. As a result of a high correlation between the glenoid asymmetry and mean moment at the glenohumeral joint, it was distinguished as a significant factor influencing shoulder strength. The application of these data lies in the greater ability to predict individual shoulder strength based on the measured area of glenoid asymmetry.

As the next step in this approach, it is proposed that each element of a joint can be calculated individually as a function of the external load and the geometry of the articulating surfaces, muscles and ligaments. Should the calculations performed for the different joints and loading reveal that the individual differences in geometry have influence on the Maximum Acceptable Workload (MAW), the current standards should be changed to reflect this.

With this approach, it is possible to reduce musculoskeletal injuries by assessing individual acceptable loads during different work activities. This method can be useful for minimizing incompatibilities between the individual worker's physical capabilities and job demands towards preventing work-related shoulder injuries.

Only through continued attention to the anatomy, by accurately describing, measuring and analyzing the geometric characteristics of the joints, may there be possible explanations of various structures and their relationships to body strength and optimized performance taking into account individual strength.

Raja Parasuraman
George Mason University

Individual Differences in Cognitive Performance Modulated by Neurotransmitter Genes

The goal of neuroergonomics is to harness the power of neuroscience to the design of human-machine systems for safety and efficiency. In addition to interface design, selection and training of operators to meet system demands represents another approach to optimizing human-machine performance. It is well known that there are large individual differences in most human information processing functions, including attention, memory, decision making, and action selection. In the past, quantification of individual differences in these functions has been dominated by the idiographic approach, in which statistical techniques are used to analyze regularities in patterns of individual performance and to link them to questionnaire assessments of personality and intelligence. In this talk I describe a new, complementary approach that capitalizes on the breakthroughs provided by the success of the Human Genome Project in 2001. This work examines how single nucleotide polymorphisms (SNPs) of neurotransmitter genes modulate different aspects of cognitive function. We find that cholinergic and dopaminergic genes differentially modulate individual differences in spatial attention and working memory, respectively (Parasuraman, Greenwood, Kumar, & Fossella, 2005). Different SNPs of genes involved in dopamine synthesis and breakdown can be linked to both spatial and verbal working memory. Specifically, the "gene dose" of a given SNP can be shown to predict, with moderate to high effect sizes, individual differences in working memory capacity. Implications for individual difference assessment, selection, and training are discussed.

Waldemar Karwowski*, Krystyna Gielo-Perczak**
* University of Louisville
** Liberty Mutual Research Institute for Safety

Modeling Individual Differences in Human Performance: Applications of the Theory of Neuro-Ecological Control Networks

Individual differences in human performance are reflected in neural processes of human thought, perception, emotion, will, memory, and imagination. Therefore, the description of human operators as individuals who actively participate in purposeful tasks and their performance on such tasks, should reflect the complexity of human brain activities, including cognition and the dynamic processes of knowing. Furthermore, the models of individual human performance should take into account human emotions, imagination and intuition in reference to affordances of the environment. In this paper we expand on our theory of *human neuro-ecological control networks* (HECON) as a framework for the study of individual differences based on dynamic human functioning afforded by environment. Affordances of the environment are interrelated with the personal human attributes in terms of neural manifestations of emotions and intuition. A process of *ecological perception and cognition* was defined as *affordation,* i.e. the human operator's understanding of an environment in terms of what the environment affords the individual and the acting upon such an understanding in order to exercise control over that environment. The individual's performance is modeled as an inter-connected control network that involves affordation, emotion and intuition. Such a network allows an individual to respond to specific affordances (or changes in affordances, i.e. disturbances) in the control of work environment. The entire neuro-ecological network responds to changing affordances (selected disturbances) by rearranging its patterns of connectivity. In this context, individual differences in performance are due to the differences in corresponding mappings from a set of affordances, emotions, and intuition to the individual's learning, adaptive- and tuning-based strategies for action.

William S. Marras
The Ohio State University

The Influence of Individual Factors on the Load-Tolerance

McGill (1997) proposed that injuries and disorders are initiated when a biomechanical load imposed on a tissue exceeds the tissue tolerance. This situation can occur when loads become excessive, as when heavy objects are lifted, or when the tolerance has decreased, such as occurs through aging or cumulative trauma. The load-tolerance relationship might also explain the casual pathways associated with the various research disciplines. It is evident that physical work factors can influence the magnitude and nature of the loading

occurring on the musculoskeletal system. However, similar processes might be at play with psychosocial factors and organizational factors (Marras, 2004). These factors could conceivably cause muscle recruitment patterns to change and increase loading on musculoskeletal tissues. In addition, tolerance may be affected by individual genetic and psychological factors as well as through previous loading history (CTDs or adaptation). Each of these factors can either lower or raise the tolerance to loading and can, therefore, influence the load-tolerance relationship. The load-tolerance relationship initiates a sequence of events relating to potential MSD pain symptom perception and reporting.

In the studies (Marras *et al.*, 2000), a direct interaction between psychosocial stress and biomechanical response of the musculoskeletal system was identified that might explain why psychosocial factors increase MSD risk. The data revealed that individuals responded in dramatically different fashion in response to the stress conditions. Individual responses were tied to individual personality characteristics that interacted strongly with psychosocial stress. In particular, those subjects who were classified as introverts via the Myers-Briggs personality inventory experienced a 14% increase in spine compression and a 27% increase in lateral spine shear under the stressed conditions compared with their extravert counterparts who experienced far less increases in spinal load (4-6%) under the stressed condition. Similarly "the intuitor" personally trait was associated with much greater spine loading (10-25%) responses under the stressed conditions compared to negligible increases in spine loading experienced by the "sensor" counterparts. Hence, this study has shown that interactions between both individual factors (personality traits) and psychosocial stress are able to explain much of the variability in subject response when performing physically demanding tasks.

In similar manner, a more recent study has been able to show how the degree of mental processing and pacing required during a physical task interact to strongly influence biomechanical loading of the spine (Davis *et.al*, 2002). In this study subjects were asked to perform lifting tasks under fast and slow pacing conditions while the mental processing requirements were altered. Mental processing consisted of high or low decision making tasks as well as varying levels of object placement complexity. Pacing by itself did not have much of an influence on spinal loading. However, when task complexity was great and high levels of mental processing were required, these factors interacted with pacing to increase spinal loading by 25-65%. Here again, the mechanism of loading was traced to significant increases in muscle coactivation.

These studies demonstrate that a rich and logical pattern of evidence of the influence of individual factors on the load-tolerance relationship if researches are wiling to reach beyond the traditional boundaries in each field of study. Most scientific advances are occurring at the intersection of scientific disciplines and interdisciplinary investigations have enriched our body of knowledge. The field of individual factors influence in performance is no different. We expect that significant leaps in knowledge will occur as we investigate the interactions between the various traditional and no-traditional disciplines that have explored individual differences.

REFERENCES

Davis KG, Marras WS, Heaney CA, Waters TR, Gupta P. (2002, The impact of mental processing and pacing on spine loading: 2002 Volvo Award in biomechanics. *Spine*, 1;27(23):2645-53

Gielo-Perczak K (2004) Maximum one-handed pull force and its relation to shoulder geometry Occupational Ergonomics 3:1-13.

Gielo-Perczak K and Karwowski W (2003)Ecological models of human performance based on affordance, emotion and intuition, *Ergonomics*, 46, 310-326.

Karwowski W and Gielo-Perczak K (2002) A theory of human ecological control networks: the role of affordation, emotion, and intuition in the design of human environments, *The Japanese Journal of Ergonomics*, 38 Suppl, 26-33.

Marras WS. (2004), State-of-the-art research perspectives on musculoskeletal disorder causation and control: the need for an intergraded understanding of risk. *J Electr.& Kinesiol.*,14(1):1-5.

Marras WS, Davis KG, Heaney CA, Maronitis AB, Allread WG. (2000), The influence of psychosocial stress, gender, and personality on mechanical loading of the lumbar spine. *Spine*, 1;25(23):3045-54.

McGill, SM. (1997), The biomechanics of low back injury: implications on current practice in industry and the clinic, *Journal of Biomechanics*, 30, 465-475.

Parasuraman R., Greenwood PM, Kumar, R, Fossella, J. (2005). Beyond heritability: Neurotransmitter genes differentially modulate visuospatial attention and working memory. *Psychological Science*, 16(3): 200-207.

Smith, K. & Liberg, M. (2005). Paintball 2: A Study of Remote-Controlled Tactile Stimulation for the Simulation of Leader Presence. *Proposal Accepted for presentation at the 49th Annual Meeting of the Human Factors and Ergonomics Society*. Orlando, FL. September 26-30.

PAINTBALL 2: A STUDY OF REMOTE-CONTROLLED VIBRO-TACTILE STIMULATION FOR THE SIMULATION OF LEADER PRESENCE

Kip Smith
Linköping Institute of Technology
Linköping, Sweden

Martin Liberg
The College of Skövde
Skövde, Sweden

At the HFES meeting in Denver we discussed experiments that documented a consistent decrement in performance associated with remote command and control (Pangburn, Freund, Pangburn, & Smith, 2003). The venues for these experiments were paintball assault lanes. Our participants were slower to obey verbal orders given by a remote leader than by a leader who was physically present. We have subsequently shown that the delay is due to the lack of leader presence and not to reliance upon mediated communication (Smith, 2004). Here we discuss an experiment that tests the hypothesis that gentle remote-controlled vibro-tactile stimulation by a remote leader can compensate for the observed decrement. The data encourage us to suggest that it would be prudent to evaluate further the effectiveness of remote-controlled vibro-tactile stimulation prior to finalizing plans for remote command and control of the dismounted infantry.

INTRODUCTION

The study reported here is the latest in a series on the impact of leader presence and remote command and control on response times to verbal commands in a simulated combat environment.

Previous Work

In previous experiments, we manipulated leader presence at three levels – a collocated leader who issued orders by yelling (the "yell" condition), a collocated leader who issued orders by radio ("radio"), and a remote leader who issued orders by radio ("remote"). The leader issued verbal orders "Move!" and "Fire!" to a participant playing the role of soldier. The data from those experiments, Figure 1, reveal that (1) participants were consistently slower to obey the commands when they were issued by a remote leader (Pangburn, Freund, Pangburn, & Smith, 2003; Smith, 2004), and (2) the delay cannot be attributed to the use of radio communication. The main effect for leader presence was significant, $F(2, 30) = 4.370$, $p < .03$, $\eta^2 = 0.047$ for response times to commands to move (more times) and $F(2, 30) = 3.353$, $p < .05$, $\eta^2 = 0.058$ for fire times.

The implication of these findings is that *leader presence matters*. Response time, a key measure of performance in any command and control environment, suffers when the leader is not physically present.

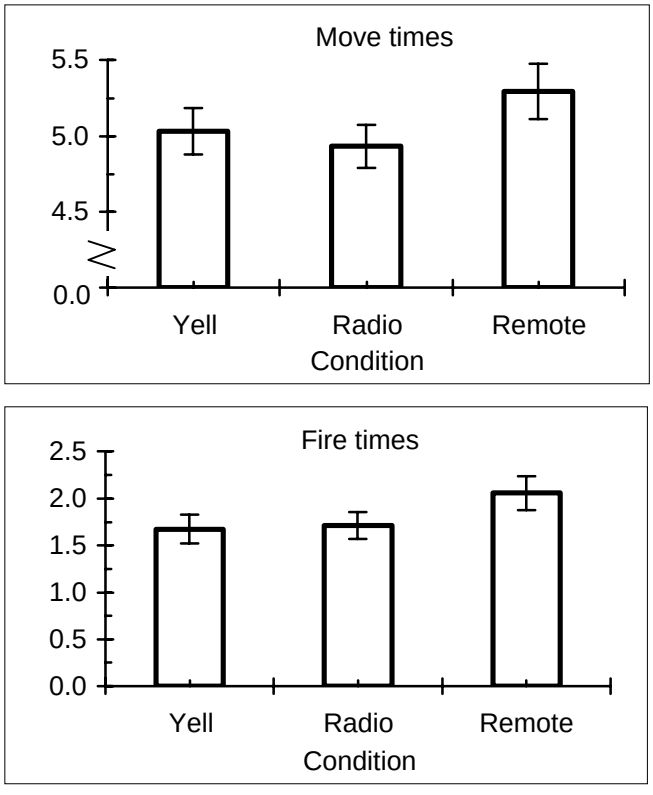


Figure 1. Summary of previous work showing the means and standard errors of response times to the verbal commands "MOVE!" (above) and "FIRE!" (below). Responses were consistently slower when the

command was issued by a remote leader (from Smith, 2004).

Is it possible to compensate for the lack of leader presence in remote command and control? Can we design a communication device that makes it "feel like" the leader is present? The purpose of the study reported here is to test whether gentle remote-controlled vibro-tactile stimulation might be a viable tool for simulating leader presence.

Network-based defense

The domain of application of our studies is a proposed redesign of the organizational structure of military units of four to eight foot soldiers known as *fire teams*. A fire team is the basic fighting unit in the dismounted infantry. The U. S. Army and the Swedish Ministry of Defense have independently started processes that will change how fire teams are formed and how they are led (Brehmer & Sundin, 2000; Friman, 2000; Hébert, 2003). The proposed redesign is called *network-based defense* in Sweden. Soldiers on the battlefield will no longer take their orders from an officer whom they can see. Instead, orders will be given by a leader located in a safe position away from the battlefield.

One consequence of this fundamental change in team dynamics is that the members of a fire team may not know their leader. A second consequence is that all communication between soldiers and officers must be mediated by radios or other portable communication devices. Accordingly, part of the vision for the infantry is to require all soldiers to carry various types of sensors and telemetry equipment so that their remote leader can see and hear what they see and hear (Hébert, 2003).

The U. S. and Swedish military are committed to effecting a transition to remote command and control in spite of the associated delay in response times to verbal commands. Given the steady advance of electronic technology in the field of battle, it is reasonable to ask whether it is possible to design a simple and lightweight technological mitigation of the decrement in cognitive performance that we have found.

Touch

Physical contact has been seen to create a basis for increased compliance to a request (Guéguen, 2004). When socially and culturally appropriate, being touched by another person has been seen to induce a positive mood in those who are touched (Henley, 1973). Hall (1996) has shown that (in western cultures) supportive touch has its greatest impact on performance when applied to the top of the back or to the upper arm. We suggest that culturally appropriate remote-controlled touch may provide a mechanism for quickly establishing virtual leader presence.

Hypothesis

If remote-controlled vibro-tactile stimulation successfully simulates leader presence, response times to verbal orders issued by a remote leader will be faster when those commands are accompanied by the simulated touch.

METHOD

Participants

A total of 24 people took part in the experiment, 22 men and 2 women. Ages ranged from 19 to 27 years with a median of 22. Eight of the men and one of the women had previous military experience lasting between 3 and 12 months. Twenty-two were Swedish, two were Icelandic. Twelve played the role of soldier in the control condition first and twelve played the role of soldier in the touch condition first in a repeated-measures design.

Experimental setting and apparatus

Layout. The experiment was conducted inside a 40 x 16 m paintball arena in Tidan, 15 km north of Skövde, Sweden. The assault lane, sketched in Figure 2, was delineated by eight large (2.5 m long by 1.5 m high) inflatable protective barriers, four on each side, 8 m apart. A small cup containing five paintballs was placed behind each barrier. At the far end of the lane stood a protective wall with two horizontal slots (0.5 m long by 0.2 m high) at shoulder height. The wall formed the sniper's hardened position. The slots allowed the sniper to shoot at (and be shot by) the soldier. A table made of shipping pallets was placed directly in front of the bunker. Five targets, 1.5 liter plastic bottles filled with colored water, four red and one blue, were placed in random order in a row on the table.

The remote leader and the observer stood in an observation room above and behind the assault lane where they could not be seen by the participant playing the role of soldier.

Tactile stimuli. A remote-controlled vibrator package was assembled by an undergraduate research assistant at the College of Skövde. It had two parts, a tactile-stimulus unit and a communications unit. The

participant playing the role of soldier wore the package in both conditions of the repeated-measures design.

The communication unit consisted of two walkie-talkies, tuned to the same frequency, and a sound-activated relay. One of the walkie-talkies was the remote leader's sending unit and the other the soldier's receiving unit. The leader initiated a tactile stimulus by pushing the 'call' button on the sending unit. This signal was received by the soldier's walkie-talkie and sent electronically to a sound-activated relay. The receiving unit and the relay were worn on a waist belt by the soldier. The relay was, in turn, connected by wire to the tactile-stimulus unit.

The tactile stimuli were delivered by a small metal box containing an electric motor driving an eccentric cam. The stimulus was a mild buzzing rumble that lasted for 2 seconds. The box was fastened by a belt that wrapped around the shoulders and ribcage so that the box sat between the shoulder blades. The brief buzzing on the back was intended to simulate tactile communication by a collocated leader.

Caveat

As noted by our reviewers, a mild buzzing rumble bears little resemblance to a pat on the back or a kick in the butt. Our intent in this study was to demonstrate the utility of the concept of remote-controlled vibro-tactile stimulation, not to replicate the experience of human contact.

Design

To test whether vibro-tactile stimulation can simulate leader presence and, as a consequence, reduce the delays associated with remote command and control, this experiment used two conditions of leader presence in a repeated-measures design. In the control condition, a remote leader issued verbal orders using a radio link; in the touch condition, the leader activated the tactile-stimulus unit at the same time as he issued an order. There were two types of orders, to move (advance to the next barrier in the lane) and to shoot (at the targets in front of the sniper's bunker).

Procedure

The experiment was conducted in Swedish. All participants read and signed an informed consent form alerting them to the dangers of paintball impacts.

The soldier's initial position was behind the barrier furthest from the sniper. After loading his or her rifle with the five paintballs from the small cup located behind the barrier, the soldier used a voice-activated head-set to inform the leader that he or she had reloaded ("laddat"). On the leader's order to fire ("eld"), the solder attempted to knock down the four "hostile" (red) targets on the table in front of the sniper's bunker without hitting the "friendly" (blue) target. When out of ammunition, the soldier informed the leader that the rifle was empty ("slut"). On the leader's order to advance ("framåt"), the soldier moved to the next protective barrier on the other side of the lane. The cycle was repeated at each of the eight barriers. The 40 paintballs, 5 at each barrier, were the soldier's only ammunition.

Dependent measures

The dependent variable was the time it took for the participant playing the role of soldier to obey an order from the remote leader. The criterion for completion differed across the two types of orders. For an order to fire, the time interval ended when the soldier first opened fire. For the order to move, the time interval ended with the soldier came to a full, controlled stop behind the next barrier.

Analyses

Two-factor mixed-effects repeated-measures ANOVA were conducted to assess the significance of the manipulation of touch and the possibility of sequence effects. The sequence of data acquisition (control-first, touch-first) was the blocking variable at two levels. The repeated measure was response time to commands issued with and without tactile stimulation.

RESULTS

Move times

The means and standard errors for the response times to orders to move ("move times") are shown in Figure 3. The open columns in Figure 3 are the data for participants who ran the assault lane in the touch condition first. Accordingly the two shorter columns on the outside are data collected in the touch condition. As expected, response times were faster in the touch condition in both sequences of data acquisition.

Neither sequence, $F(1, 22) = 0.027$, $\underline{p} > .86$, nor the interaction of participant-within-sequence and condition, $F(1, 22) = 1.779$, $\underline{p} > .19$, was significant. As expected, touch was significant, $F(1, 22) = 47.132$, $\underline{p} < .001$, $\eta^2 = .20$. This result suggests that verbal orders to move across an assault lane where a sniper is

maintaining live paintball fire are obeyed faster when accompanied by a tactile stimulus.

Fire times

Figure 4 presents the means and standard errors for response times to the command to fire ("fire times"). Once again, the open columns are the data for participants who ran the assault lane in the touch condition first and the outer columns are data collected in the touch condition. Response times were virtually identical in all but the control-first trials. Participants who experienced the control condition first were markedly faster in the touch condition of the second trial. In contrast, those who experienced the touch condition first did not slow down in the second trial. This pattern of data suggests there may be differential transfer of the prior experience of tactile stimulus.

The repeated-measures ANOVA indicates that the effect of sequence is not significant, $F(1, 22) = 1.888$, $p > .18$, but that the interaction of participant-within-sequence and experimental condition was significant, $F(1, 22) = 11.718$, $p < .003$. This interaction is evident in the pattern of means shown in Figure 4. This evidence for differential transfer, while unexpected, is auspiciously in the desired direction. The manipulation of touch was again highly significant, $F(1, 22) = 13.102$, $p < .002$, $\eta^2 = .15$. This result suggests that verbal orders to fire in the face of live paintball fire are obeyed faster when accompanied by a tactile stimulus.

DISCUSSION

The expected results were obtained for both types of orders. Orders to move and to shoot issued by a remote leader were obeyed more quickly when accompanied by a tactile stimulus. The effect of the tactile stimulus was more pronounced for orders to move. The data strongly support the contention that the administration of a mild buzzing rumble between the shoulder blades had a strong impact on participants' performance under conditions emulating live fire and remote command and control.

The analyses support the hypothesis that the administration of tactile stimuli can speed performance at obeying orders under conditions of remote command and control. The suggestion of differential transfer in the results for fire times is also encouraging. The possibility that prior experience with tactile stimuli may have lasting benefits lends further support to the claim that remote-controlled, simulated touch may become an effective tool for compensating, in part, for the hesitancy to obey remote leaders.

There are at least two competing explanations for this result. The first invokes trust as a mediating variable (Handy, 1995). The second argues for attentional capture by a subcortical, tactile pathway (e.g., Damasio, 1999). The former posits a mechanism that might well be accessible during retrospection whereas the latter posits a pre-attentive mechanism that is likely opaque to conscious awareness. We are currently conducting studies and gathering questionnaire data in the attempt to discriminate between these alternatives.

Regardless of the mechanism by which vibro-tactile stimulation simulates leader presence, ergonomic issues remain: how to apply the tactile stimuli so that they have the best possible effect? For example, it remains to be seen whether a tactile stimulus between the shoulder blades is the clearest experience of simulated touch and whether other stimuli may be more effective than a mild buzzing rumble.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Prof. Peter A. Hancock is the lead PI. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This paper benefited from comments by three anonymous HFES reviewers.

We thank the many people who have helped make this experiment possible. A partial list includes Berndt Brehmer, Monica Clement, Peter Hancock, Paul Hemeren, Jens Hilmersson, Magnus Jacobsson, Keith Pangburn, and the members of Spelapaintball, Tidan, Västergötland, Sverige.

REFERENCES

Brehmer, B. & Sundin, C. (2000). Command and control in network-centric warfare. In C. Sundin & H. Friman, (Eds.), ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6 (pp. 45-56). Stockholm: The Swedish National Defense College.

Damasio, A. R. (1999). The feeling of what happens: Body and emotion in the making of consciousness. New York: Harcourt Brace & Co.

Eden, D. (1986). Team development: Quasi-experimental confirmation among combat companies. Group and organization studies, 11(3), 133-146.

Smith, K. & Liberg, M. (2005). Paintball 2: A Study of Remote-Controlled Tactile Stimulation for the Simulation of Leader Presence. *Proposal Accepted for presentation at the 49th Annual Meeting of the Human Factors and Ergonomics Society*. Orlando, FL. September 26-30.

Friman, H. (2000). Decision making in a ROLF staff environment. In C. Sundin & H. Friman, (Eds.), ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6 (pp. 67-74). Stockholm: The Swedish National Defense College.

Guéguen, N. (2004). Nonverbal encouragement of participation in a course: The effect of touching. Social Psychology of Education, 7(1), 89-98.

Hall, J. (1996). Touch, status, and gender at professional meetings. Journal of Nonverbal Behavior, 20, 23-44.

Handy, C. (1995). Trust and the virtual organization. Harvard Business Review, 73, 40-50.

Hébert, R. (2003). Psychological science and the transformation of the military. American Psychological Society Observer, 16(9).

Henley, N. (1973). The politics of touch. In P. Brown (Ed.), Radical psychology. New York: Harper & Row.

Pangburn, K., Freund, S., Pangburn, H., & Smith, K. (2003). Beyond Milgram: An experimental study of leader presence. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society, 364-368.

Smith, K. (2004). The adverse impact of remote command and control under live fire. Human Performance in Extreme Environments.

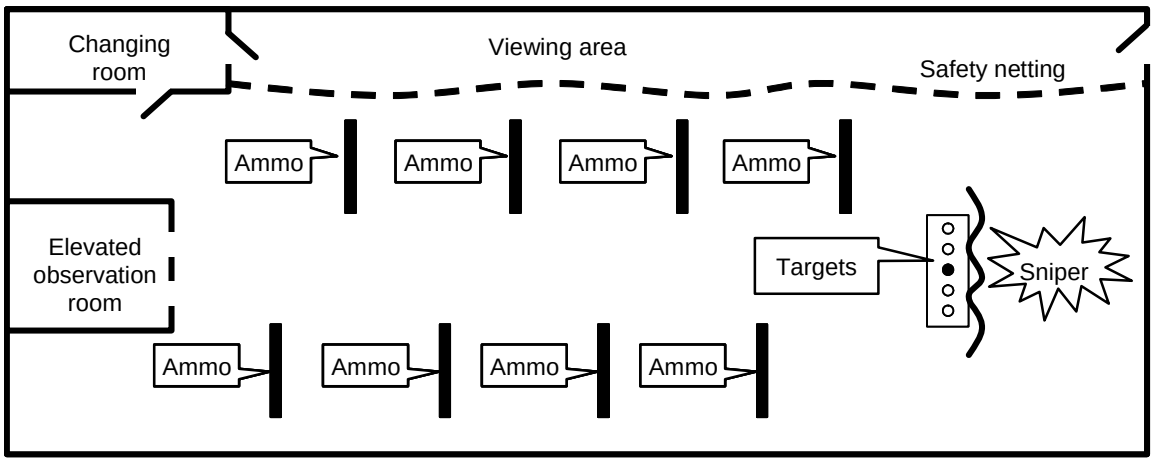


Figure 2. Layout of the paintball assault lane. Participants in the role of soldier started at the barrier closest to the observation room and, on command, advanced from barrier to barrier toward the sniper to obtain ammunition with which they, on command, shot at the targets.

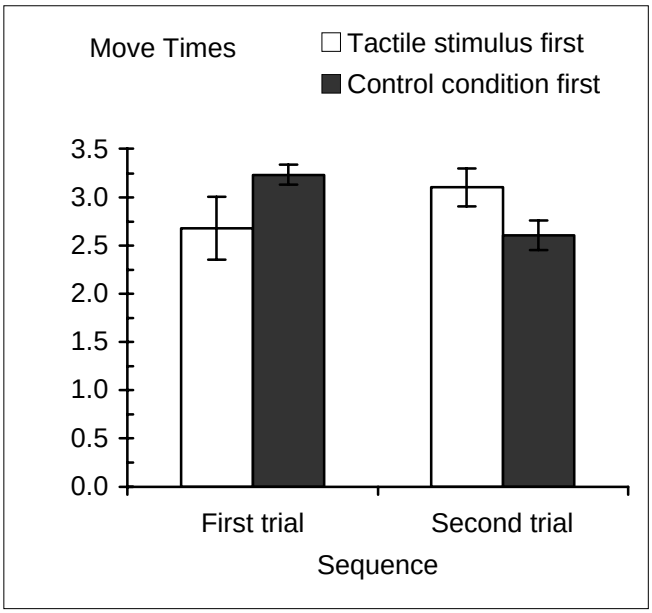


Figure 3. Means and standard errors of response times to the verbal command "MOVE!" showing faster responses when the command was accompanied by gentle vibro-tactile stimulation.

Smith, K. & Liberg, M. (2005). Paintball 2: A Study of Remote-Controlled Tactile Stimulation for the Simulation of Leader Presence. *Proposal Accepted for presentation at the 49th Annual Meeting of the Human Factors and Ergonomics Society*. Orlando, FL. September 26-30.

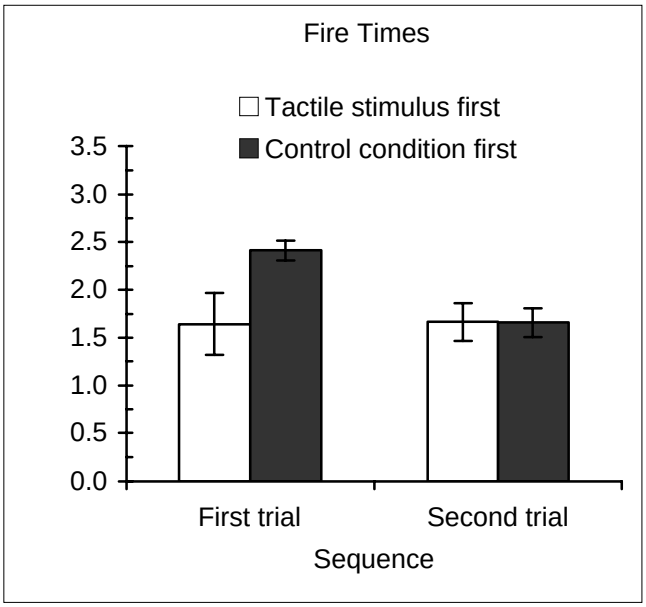


Figure 4. Means and standard errors of response times to the verbal command "FIRE!" showing faster responses when the command was accompanied or preceded by gentle vibro-tactile stimulation.

Smith, K. & Liberg, M. (2005). Paintball 2: A Study of Remote-Controlled Tactile Stimulation for the Simulation of Leader Presence. *Proposal Accepted for presentation at the 49th Annual Meeting of the Human Factors and Ergonomics Society*. Orlando, FL. September 26-30.

PAINTBALL 2: A STUDY OF REMOTE-CONTROLLED VIBRO-TACTILE STIMULATION FOR THE SIMULATION OF LEADER PRESENCE

Kip Smith
Linköping Institute of Technology
Linköping, Sweden

Martin Liberg
The College of Skövde
Skövde, Sweden

At the HFES meeting in Denver we discussed experiments that documented a consistent decrement in performance associated with remote command and control (Pangburn, Freund, Pangburn, & Smith, 2003). The venues for these experiments were paintball assault lanes. Our participants were slower to obey verbal orders given by a remote leader than by a leader who was physically present. We have subsequently shown that the delay is due to the lack of leader presence and not to reliance upon mediated communication (Smith, 2004). Here we discuss an experiment that tests the hypothesis that gentle remote-controlled vibro-tactile stimulation by a remote leader can compensate for the observed decrement. The data encourage us to suggest that it would be prudent to evaluate further the effectiveness of remote-controlled vibro-tactile stimulation prior to finalizing plans for remote command and control of the dismounted infantry.

INTRODUCTION

The study reported here is the latest in a series on the impact of leader presence and remote command and control on response times to verbal commands in a simulated combat environment.

Previous Work

In previous experiments, we manipulated leader presence at three levels – a collocated leader who issued orders by yelling (the "yell" condition), a collocated leader who issued orders by radio ("radio"), and a remote leader who issued orders by radio ("remote"). The leader issued verbal orders "Move!" and "Fire!" to a participant playing the role of soldier. The data from those experiments, Figure 1, reveal that (1) participants were consistently slower to obey the commands when they were issued by a remote leader (Pangburn, Freund, Pangburn, & Smith, 2003; Smith, 2004), and (2) the delay cannot be attributed to the use of radio communication. The main effect for leader presence was significant, $F(2, 30) = 4.370$, $p < .03$, $\eta^2 = 0.047$ for response times to commands to move (more times) and $F(2, 30) = 3.353$, $p < .05$, $\eta^2 = 0.058$ for fire times.

The implication of these findings is that *leader presence matters*. Response time, a key measure of performance in any command and control environment, suffers when the leader is not physically present.

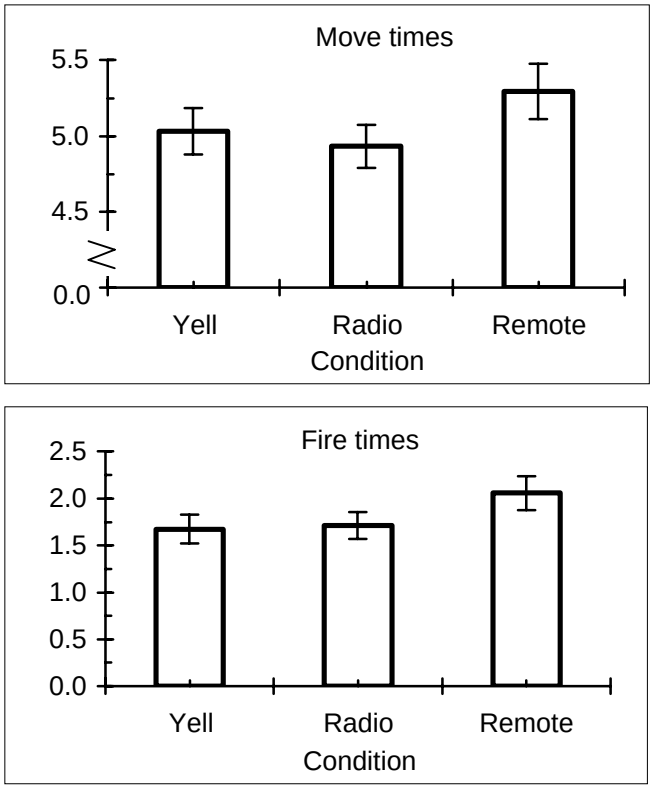


Figure 1. Summary of previous work showing the means and standard errors of response times to the verbal commands "MOVE!" (above) and "FIRE!" (below). Responses were consistently slower when the

command was issued by a remote leader (from Smith, 2004).

Is it possible to compensate for the lack of leader presence in remote command and control? Can we design a communication device that makes it "feel like" the leader is present? The purpose of the study reported here is to test whether gentle remote-controlled vibro-tactile stimulation might be a viable tool for simulating leader presence.

Network-based defense

The domain of application of our studies is a proposed redesign of the organizational structure of military units of four to eight foot soldiers known as *fire teams*. A fire team is the basic fighting unit in the dismounted infantry. The U. S. Army and the Swedish Ministry of Defense have independently started processes that will change how fire teams are formed and how they are led (Brehmer & Sundin, 2000; Friman, 2000; Hébert, 2003). The proposed redesign is called *network-based defense* in Sweden. Soldiers on the battlefield will no longer take their orders from an officer whom they can see. Instead, orders will be given by a leader located in a safe position away from the battlefield.

One consequence of this fundamental change in team dynamics is that the members of a fire team may not know their leader. A second consequence is that all communication between soldiers and officers must be mediated by radios or other portable communication devices. Accordingly, part of the vision for the infantry is to require all soldiers to carry various types of sensors and telemetry equipment so that their remote leader can see and hear what they see and hear (Hébert, 2003).

The U. S. and Swedish military are committed to effecting a transition to remote command and control in spite of the associated delay in response times to verbal commands. Given the steady advance of electronic technology in the field of battle, it is reasonable to ask whether it is possible to design a simple and lightweight technological mitigation of the decrement in cognitive performance that we have found.

Touch

Physical contact has been seen to create a basis for increased compliance to a request (Guéguen, 2004). When socially and culturally appropriate, being touched by another person has been seen to induce a positive mood in those who are touched (Henley, 1973). Hall (1996) has shown that (in western cultures) supportive touch has its greatest impact on performance when applied to the top of the back or to the upper arm. We suggest that culturally appropriate remote-controlled touch may provide a mechanism for quickly establishing virtual leader presence.

Hypothesis

If remote-controlled vibro-tactile stimulation successfully simulates leader presence, response times to verbal orders issued by a remote leader will be faster when those commands are accompanied by the simulated touch.

METHOD

Participants

A total of 24 people took part in the experiment, 22 men and 2 women. Ages ranged from 19 to 27 years with a median of 22. Eight of the men and one of the women had previous military experience lasting between 3 and 12 months. Twenty-two were Swedish, two were Icelandic. Twelve played the role of soldier in the control condition first and twelve played the role of soldier in the touch condition first in a repeated-measures design.

Experimental setting and apparatus

Layout. The experiment was conducted inside a 40 x 16 m paintball arena in Tidan, 15 km north of Skövde, Sweden. The assault lane, sketched in Figure 2, was delineated by eight large (2.5 m long by 1.5 m high) inflatable protective barriers, four on each side, 8 m apart. A small cup containing five paintballs was placed behind each barrier. At the far end of the lane stood a protective wall with two horizontal slots (0.5 m long by 0.2 m high) at shoulder height. The wall formed the sniper's hardened position. The slots allowed the sniper to shoot at (and be shot by) the soldier. A table made of shipping pallets was placed directly in front of the bunker. Five targets, 1.5 liter plastic bottles filled with colored water, four red and one blue, were placed in random order in a row on the table.

The remote leader and the observer stood in an observation room above and behind the assault lane where they could not be seen by the participant playing the role of soldier.

Tactile stimuli. A remote-controlled vibrator package was assembled by an undergraduate research assistant at the College of Skövde. It had two parts, a tactile-stimulus unit and a communications unit. The

participant playing the role of soldier wore the package in both conditions of the repeated-measures design.

The communication unit consisted of two walkie-talkies, tuned to the same frequency, and a sound-activated relay. One of the walkie-talkies was the remote leader's sending unit and the other the soldier's receiving unit. The leader initiated a tactile stimulus by pushing the 'call' button on the sending unit. This signal was received by the soldier's walkie-talkie and sent electronically to a sound-activated relay. The receiving unit and the relay were worn on a waist belt by the soldier. The relay was, in turn, connected by wire to the tactile-stimulus unit.

The tactile stimuli were delivered by a small metal box containing an electric motor driving an eccentric cam. The stimulus was a mild buzzing rumble that lasted for 2 seconds. The box was fastened by a belt that wrapped around the shoulders and ribcage so that the box sat between the shoulder blades. The brief buzzing on the back was intended to simulate tactile communication by a collocated leader.

Caveat

As noted by our reviewers, a mild buzzing rumble bears little resemblance to a pat on the back or a kick in the butt. Our intent in this study was to demonstrate the utility of the concept of remote-controlled vibro-tactile stimulation, not to replicate the experience of human contact.

Design

To test whether vibro-tactile stimulation can simulate leader presence and, as a consequence, reduce the delays associated with remote command and control, this experiment used two conditions of leader presence in a repeated-measures design. In the control condition, a remote leader issued verbal orders using a radio link; in the touch condition, the leader activated the tactile-stimulus unit at the same time as he issued an order. There were two types of orders, to move (advance to the next barrier in the lane) and to shoot (at the targets in front of the sniper's bunker).

Procedure

The experiment was conducted in Swedish. All participants read and signed an informed consent form alerting them to the dangers of paintball impacts.

The soldier's initial position was behind the barrier furthest from the sniper. After loading his or her rifle with the five paintballs from the small cup located behind the barrier, the soldier used a voice-activated head-set to inform the leader that he or she had reloaded ("laddat"). On the leader's order to fire ("eld"), the solder attempted to knock down the four "hostile" (red) targets on the table in front of the sniper's bunker without hitting the "friendly" (blue) target. When out of ammunition, the soldier informed the leader that the rifle was empty ("slut"). On the leader's order to advance ("framåt"), the soldier moved to the next protective barrier on the other side of the lane. The cycle was repeated at each of the eight barriers. The 40 paintballs, 5 at each barrier, were the soldier's only ammunition.

Dependent measures

The dependent variable was the time it took for the participant playing the role of soldier to obey an order from the remote leader. The criterion for completion differed across the two types of orders. For an order to fire, the time interval ended when the soldier first opened fire. For the order to move, the time interval ended with the soldier came to a full, controlled stop behind the next barrier.

Analyses

Two-factor mixed-effects repeated-measures ANOVA were conducted to assess the significance of the manipulation of touch and the possibility of sequence effects. The sequence of data acquisition (control-first, touch-first) was the blocking variable at two levels. The repeated measure was response time to commands issued with and without tactile stimulation.

RESULTS

Move times

The means and standard errors for the response times to orders to move ("move times") are shown in Figure 3. The open columns in Figure 3 are the data for participants who ran the assault lane in the touch condition first. Accordingly the two shorter columns on the outside are data collected in the touch condition. As expected, response times were faster in the touch condition in both sequences of data acquisition.

Neither sequence, $F(1, 22) = 0.027$, $\underline{p} > .86$, nor the interaction of participant-within-sequence and condition, $F(1, 22) = 1.779$, $\underline{p} > .19$, was significant. As expected, touch was significant, $F(1, 22) = 47.132$, $\underline{p} < .001$, $\eta^2 = .20$. This result suggests that verbal orders to move across an assault lane where a sniper is

maintaining live paintball fire are obeyed faster when accompanied by a tactile stimulus.

Fire times

Figure 4 presents the means and standard errors for response times to the command to fire ("fire times"). Once again, the open columns are the data for participants who ran the assault lane in the touch condition first and the outer columns are data collected in the touch condition. Response times were virtually identical in all but the control-first trials. Participants who experienced the control condition first were markedly faster in the touch condition of the second trial. In contrast, those who experienced the touch condition first did not slow down in the second trial. This pattern of data suggests there may be differential transfer of the prior experience of tactile stimulus.

The repeated-measures ANOVA indicates that the effect of sequence is not significant, $F(1, 22) = 1.888$, $p > .18$, but that the interaction of participant-within-sequence and experimental condition was significant, $F(1, 22) = 11.718$, $p < .003$. This interaction is evident in the pattern of means shown in Figure 4. This evidence for differential transfer, while unexpected, is auspiciously in the desired direction. The manipulation of touch was again highly significant, $F(1, 22) = 13.102$, $p < .002$, $\eta^2 = .15$. This result suggests that verbal orders to fire in the face of live paintball fire are obeyed faster when accompanied by a tactile stimulus.

DISCUSSION

The expected results were obtained for both types of orders. Orders to move and to shoot issued by a remote leader were obeyed more quickly when accompanied by a tactile stimulus. The effect of the tactile stimulus was more pronounced for orders to move. The data strongly support the contention that the administration of a mild buzzing rumble between the shoulder blades had a strong impact on participants' performance under conditions emulating live fire and remote command and control.

The analyses support the hypothesis that the administration of tactile stimuli can speed performance at obeying orders under conditions of remote command and control. The suggestion of differential transfer in the results for fire times is also encouraging. The possibility that prior experience with tactile stimuli may have lasting benefits lends further support to the claim that remote-controlled, simulated touch may become an effective tool for compensating, in part, for the hesitancy to obey remote leaders.

There are at least two competing explanations for this result. The first invokes trust as a mediating variable (Handy, 1995). The second argues for attentional capture by a subcortical, tactile pathway (e.g., Damasio, 1999). The former posits a mechanism that might well be accessible during retrospection whereas the latter posits a pre-attentive mechanism that is likely opaque to conscious awareness. We are currently conducting studies and gathering questionnaire data in the attempt to discriminate between these alternatives.

Regardless of the mechanism by which vibro-tactile stimulation simulates leader presence, ergonomic issues remain: how to apply the tactile stimuli so that they have the best possible effect? For example, it remains to be seen whether a tactile stimulus between the shoulder blades is the clearest experience of simulated touch and whether other stimuli may be more effective than a mild buzzing rumble.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Prof. Peter A. Hancock is the lead PI. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This paper benefited from comments by three anonymous HFES reviewers.

We thank the many people who have helped make this experiment possible. A partial list includes Berndt Brehmer, Monica Clement, Peter Hancock, Paul Hemeren, Jens Hilmersson, Magnus Jacobsson, Keith Pangburn, and the members of Spelapaintball, Tidan, Västergötland, Sverige.

REFERENCES

Brehmer, B. & Sundin, C. (2000). Command and control in network-centric warfare. In C. Sundin & H. Friman, (Eds.), ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6 (pp. 45-56). Stockholm: The Swedish National Defense College.

Damasio, A. R. (1999). The feeling of what happens: Body and emotion in the making of consciousness. New York: Harcourt Brace & Co.

Eden, D. (1986). Team development: Quasi-experimental confirmation among combat companies. Group and organization studies, 11(3), 133-146.

Smith, K. & Liberg, M. (2005). Paintball 2: A Study of Remote-Controlled Tactile Stimulation for the Simulation of Leader Presence. *Proposal Accepted for presentation at the 49th Annual Meeting of the Human Factors and Ergonomics Society*. Orlando, FL. September 26-30.

Friman, H. (2000). Decision making in a ROLF staff environment. In C. Sundin & H. Friman, (Eds.), ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6 (pp. 67-74). Stockholm: The Swedish National Defense College.

Guéguen, N. (2004). Nonverbal encouragement of participation in a course: The effect of touching. Social Psychology of Education, 7(1), 89-98.

Hall, J. (1996). Touch, status, and gender at professional meetings. Journal of Nonverbal Behavior, 20, 23-44.

Handy, C. (1995). Trust and the virtual organization. Harvard Business Review, 73, 40-50.

Hébert, R. (2003). Psychological science and the transformation of the military. American Psychological Society Observer, 16(9).

Henley, N. (1973). The politics of touch. In P. Brown (Ed.), Radical psychology. New York: Harper & Row.

Pangburn, K., Freund, S., Pangburn, H., & Smith, K. (2003). Beyond Milgram: An experimental study of leader presence. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society, 364-368.

Smith, K. (2004). The adverse impact of remote command and control under live fire. Human Performance in Extreme Environments.

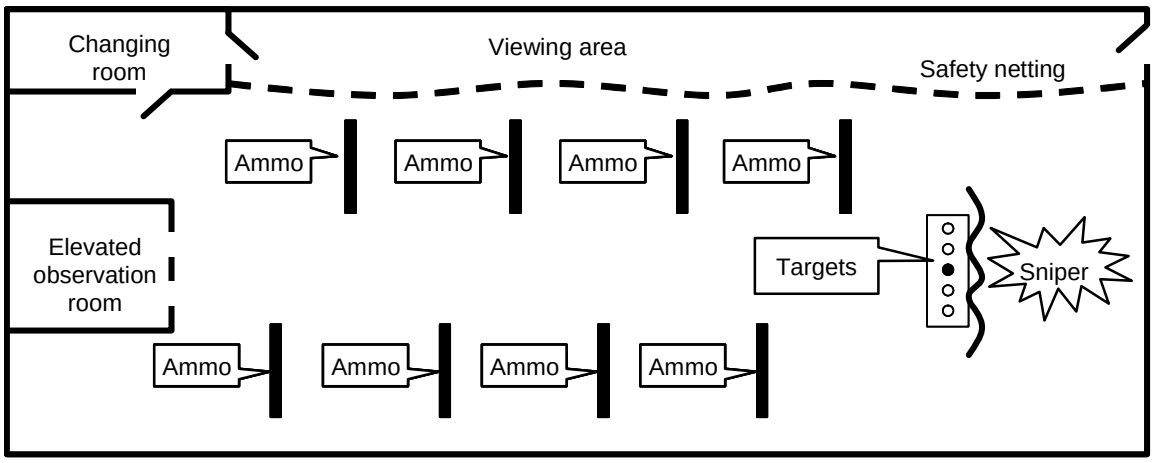


Figure 2. Layout of the paintball assault lane. Participants in the role of soldier started at the barrier closest to the observation room and, on command, advanced from barrier to barrier toward the sniper to obtain ammunition with which they, on command, shot at the targets.

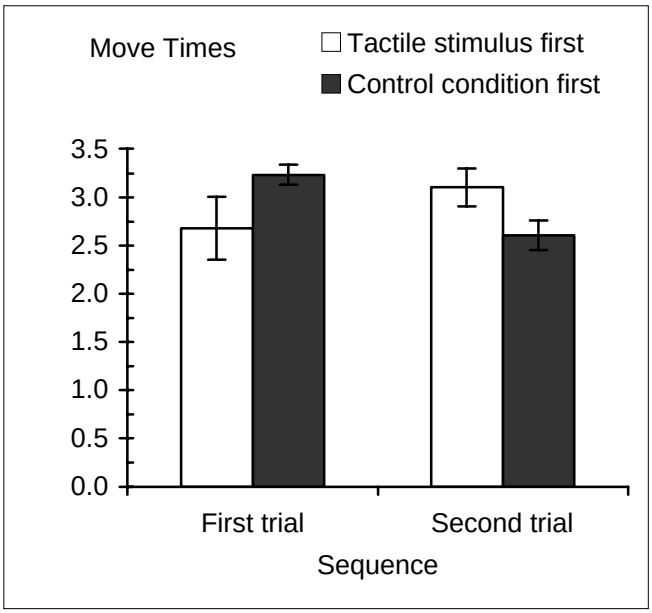


Figure 3. Means and standard errors of response times to the verbal command "MOVE!" showing faster responses when the command was accompanied by gentle vibro-tactile stimulation.

Smith, K. & Liberg, M. (2005). Paintball 2: A Study of Remote-Controlled Tactile Stimulation for the Simulation of Leader Presence. *Proposal Accepted for presentation at the 49th Annual Meeting of the Human Factors and Ergonomics Society*. Orlando, FL. September 26-30.

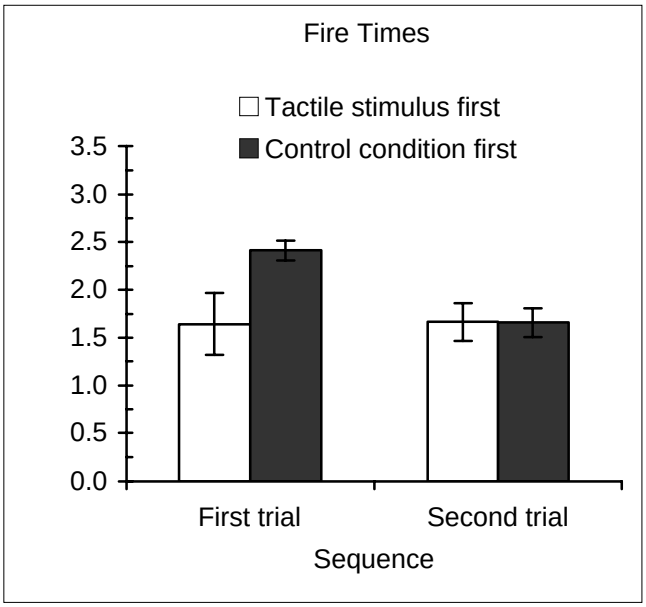


Figure 4. Means and standard errors of response times to the verbal command "FIRE!" showing faster responses when the command was accompanied or preceded by gentle vibro-tactile stimulation.

Running head: HUD USE AND LANDING PERFORMANCE

HUD use improves landing performance in restricted visibility

Örjan Goteman

Scandinavian Airlines

Kip Smith and Sidney Dekker

Linköping Institute of Technology

Abstract

The operational community has assumed that using a head-up display (HUD) instead of conventional head-down displays will increase accuracy and safety during approach and landing. This paper reports a pair of experiments that test that hypothesis. Using a Boeing 737-700 full-motion simulator flown by commercial airline pilots, we explored the effects of (1) HUD use, (2) ambient visibility, and (3) length of approach lighting on the size and location of the touchdown footprint. HUD use reduced the width of the touchdown footprint in all tested visibility and lighting conditions including visibility below the minimum allowed. HUD use had no effect on the length of the touchdown footprint. Based on these empirical data, the minimum visibility for approaches using HUD could be set lower than for approaches without a HUD.

HUD use improves landing performance in restricted visibility

This paper reports two experiments designed to assess the impact of commercial airline pilots' use of a head-up display (HUD) on approach and landing performance under reduced runway visual range (RVR) conditions. Our goal is to report experimental work designed to inform the development or revision of regulations on HUD use in low RVR conditions. We begin by reviewing previous research on HUD use and by describing the effect of how ambient RVR affects approach and landing operations. The review motivates a pair of hypotheses regarding the impact of HUD use on the size of the landing footprint. The experiments and their findings are discussed in turn with particular emphasis on the mechanisms behind the effects and the practical implications for approach and landing in restricted visibility.

HUD use in approach and landing operations

The HUD projects computer-generated aircraft flight-path and energy symbols onto a transparent screen in the pilot's primary view. Pilots can concurrently view the HUD symbology and the surrounding world. Commercially available HUDs usually replicate the information made available by the pilot's conventional flight instruments, showing aircraft attitude, speed, altitude and heading. In addition, HUDs typically contain a flight path symbol. Conformal HUD displays preserve a 1:1 relationship between the visual angles of displayed symbols and their real world referents. A conformal HUD with a flight path symbol can thus instantaneously show the pilot where the aircraft is going relative to the surrounding world. In contrast, a pilot relying on conventional flight instruments must infer the aircraft's flight path based solely on his/her perception of optical flow (Gibson, 1986).

Comparisons between HUD and head-down displays in manual flight have found that conformal HUD use improved track, speed, and altitude maintenance (Lauber et al., 1982; Martin-Emerson & Wickens, 1997). The civil operational community assumed that these HUD performance advantages over conventional head-down instrumentation could reduce the number of approach and landing incidents and accidents (Flight Safety Foundation,). Specifically, the commercial operational community assumed that HUD use could mitigate two well-documented problems associated with approach and landing: (1) visual approaches to runways without radio navigation aids or with unreliable navigation aids, and (2) the transition from instrument to external visual cues for landing in low visibility (e.g., Newman, 1995). The presumed sources of the advantage in flight-tracking performance for the HUD are that it eliminates the need for the pilot to move his or her gaze from head-down instruments to the outside world to look for maneuvering cues (Stuart, McAnally, & Meehan, 2003) and minimizes scanning requirements (Martin-Emerson and Wickens, 1997). One consequence of the transition from heads-down to the outside world is the need for visual accommodation (e.g., the visual depth of field changes from less than a meter to infinity). Because conformal HUD symbology is focused at infinity, HUD use eliminates the need for and time demand of visual accommodation, simplifying the pilot's task. The elimination of need to re-accommodate is not undisputed. Iavecchia, Iavecchia, and Roscoe (1988) reported observing over-accommodation (when the pilot's focus turns to the HUD symbology) despite the collimated optics when pilots used a HUD. This lead them to predict poorer touchdown control. Specifically, they predicted that over-accommodation would contribute to longer and shallower landings with HUD.

A series of experiments have used the ‘unusual event’ paradigm in the form of another aircraft unexpectedly taxiing onto the runway during approach to document decrements in performance associated with HUD use during landing operations (Fisher, Hanes and Price, 1980, Weintraub, Haines &Randle, 1985, Wickens and Long, 1995)). These decrements, known alternately as "attentional capture" "or cognitive tunnelling", while theoretically interesting, are not operationally relevant in low visibility operations. At a typical approach speed of 140 knots for a large airliner, the maximum available time from visual detection of an obstacle on the runway until the aircraft is passing the threshold is only around 7 seconds in RVR of 550 meters. Consequently, the responsibility for air traffic separation lies firmly with Air Traffic Control during instrument flight conditions (ICAO, 2001a). For a review of attentional capture and HUD use see Stuart, McAnally and Meehan (2003).

In this paper we focus on pilot performance during landing in approach operations under low RVR conditions. We seek to ascertain and quantify whether the HUD-related advantages for flight-path tracking during cruise established by earlier research is carried over to the landing. Data exist for certification trials for HUDs currently used in civil aviation. However, those trials used HUDs in a low visibility mode with guidance down to and including the landing flare, where there was no need for external visual cues at all (Joint Aviation Authorities [JAA], 2004a). Such data do not export well to touchdown performance using HUD in category I flight operations where obstacle clearance below the decision height is maintained by visual maneuvering. We know of no other studies with commercial pilots in high fidelity simulations that have tested whether HUD use is associated with performance advantages related to landing in restricted visibility operations predicated on visual cues for maneuvering. .

Approach operations in reduced visibility

The majority of approaches (the segment of flight immediately before touchdown) performed in US and European civil air transport operation are conducted as Category I Instrument Landing System (ILS) approaches. For a category I approach the two critical factors when making the decision of whether or not to land are the decision height and the runway visual range (RVR). Decision height refers to the aircraft's vertical distance above the runway threshold where the pilot must make the decision to land or make a go-around. As a minimum the pilot shall be able to see some elements of the approach or runway lights. The approach lights will then guide the pilot to the runway.

The RVR is a measure of horizontal visibility defined by the number of visible approach and runway lights in the ambient atmospheric conditions. If the RVR is too low at the decision height, the pilot will not be able to see any of the approach lights and must make a go-around. Although the resulting missed approach is part of normal operations (and formal procedures), it adds an undesired additional risk (ICAO, 1993).

The length of the approach lights differ from runway to runway. Geographical constraints may make the standard full length of 720 m (Full Facilities) impossible to achieve. Less length means less guidance and later contact with the approach lights in an approach. Consequently the length of the approach lights is reflected in existing regulations. For example; to commence a Category I ILS approach to a runway equipped with 720 m length of approach lights, the RVR measured at the runway must not be less than 550 m (Federal Aviation Authority [FAA], 2002, JAA 2004b). This RVR is expected to allow the pilot to see a visual segment of the ground that

contains enough of the runway approach lights to judge the aircraft's lateral position, cross track velocity, and position in roll when the aircraft is at decision height). This condition is sketched in Figure 1. The same condition viewed from the pilots position in cockpit is sketched in figure 2 (A). The figure visualizes how the visible row of Full Facility (720 m) approach lights appear from under the aircraft's nose and stretch forward towards the runway until they are obscured by the limited visibility. In figure 2 (B) we have sketched the pilot's view of the same approach lights length in lower-than-standard (450 m) RVR. Here the shorter ground visual segment is seen as fewer approach lights are available for guidance towards the runway.

For shorter approach lights lengths there is less guidance. In figure 3(A) we have sketched what the pilot sees at decision heights for approach lights with 420 meters length in standard (700 m) RVR. In figure 3(B) the pilot view of the same approach lights is shown in lower-than-standard (550 m) RVR. In all of these approach lights lengths/RVR combination the pilot does not see the runway threshold at decision height. Later, as the aircraft comes closer to the runway threshold, the pilot sees and aims at the designated touchdown point, normally 305 m down the runway

The issue of RVR and HUD use

Large commercial airlines such as American Airlines, All Nippon Airways, Delta, Scandinavian Airlines and Southwest Airlines are introducing HUDs to the flight deck at an increasing rate, making the effect of HUD use on touchdown performance an operationally relevant question. Generalizing from earlier research (Hoerner, 1979, Martin-Emerson and Wickens, 1997), HUD use with conformal symbology should improve tracking performance during approach and landing in restricted visibility. Showing the instantaneous flight path should enable the pilot to control the aircraft better below decision height and this should reduce

the variance of the touchdown point. A reduction in touchdown variability could allow approach operations using HUD in lower than standard RVR conditions. The current operating minima were not set bearing HUD operations in mind (FAA 2002, JAA 2004b) and may be too restrictive for operations using HUD, a fact that the existing legislative text acknowledges (JAA, 2004b). However, allowing a too low RVR to commence approach for HUD operations will likely reduce approach success rate. Although the focus of this paper is not on the issue of cognitive tunneling, we believe that this phenomenon could play a role at decision height. There may be a risk that the pilots will focus on the HUD symbology at decision height at the expense of the outside cues to the extent that they might not perceive the visible approach lights and thereby initiate a go-around when the approach could actually have been continued. This is not desirable from an operational standpoint. Squeezed between operational pressure for lower minima for HUD's presumed advantage and possible negative effects of HUD, regulators are being pressed to generate rules that are insufficiently grounded in research.

In light of the increasing use of HUDs in commercial aviation and the paucity of the existing knowledge base, we studied whether the established improvement in tracking performance when using a HUD would be reflected in a smaller touchdown footprint on the runway in standard and lower-than-standard RVR conditions. As there is less guidance with shorter approach lights we investigated the effect of HUD on touchdown performance for two different approach lights conditions as defined in the European legislations; Full Approach Light facilities (>720 m) and Intermediate Approach Light facilities (420-719 m) (JAA, 2004b). See table 1 for the various RVR and approach lights lengths conditions for the two experiments. We predicted that the HUD would reduce the size of the footprint both longitudinally and laterally

and that a HUD in itself would not reduce the approach rate at the RVR's used in the experiments.

Experiment 1

. In the first experiment, we assessed HUD use in visibilities at and below the required minimum RVR at simulated facilities with a system of approach lights of full length (more than 720 m). In the second experiment, we assessed HUD use in visibilities at and below the required minimum RVR at simulated facilities with a system of approach lights of intermediate length (420 or 600

Method

Participants. Forty-three pilots from a major European airline volunteered to participate. All were qualified to fly the B737-700 aircraft and had completed their HUD training for the operator. The training sessions consisted of one day of theory and two simulator sessions of four hours duration each. Experience on the B737-700 varied from 50 to more than 1000 hours. There is every reason to believe that these participants are representative of the population of commercial pilots to whom regulators need the data to generalize.

Apparatus. We used a CAE B737-700 training simulator with aircraft aerodynamics and visual angles valid for B737-700 to collect data. This six-axis full-motion simulator is approved for low-visibility operations down to an RVR of 200 m. The pilot had a field of vision of 180°/40° collimated to a focal distance greater than 10 m.

The HUD installed in the simulator was a Rockwell-Collins Flight Dynamics HGS-4000® (Head-up Guidance System), certified for low visibility operations down to and including an RVR of 200 m. The HUD symbology used in the study is shown in Figure 4. The symbology

and functionality met the production-line standard specification for the mode customarily used (IMC, instrumental meteorological condition mode) when conducting Category I instrument landing system (ILS) and non-precision approaches.

The HUD provided conformal display of flight path and flight-path guidance in the form of a ring inside the flight-path symbol. Flight-path guidance is not available in the conventional head-down instrumentation of the B737-NG with a primary flight display.

The radio navigation facilities simulated in the study conformed to the International Civil Aviation Organization (ICAO, 1996) standard for ILS radio navigation aids. Such a system transmits a beam for both vertical and lateral reference. The integrity of the transmitted beam is guaranteed to keep the aircraft within allowable airspace, safe from obstacles down to the decision height. Approaches in Experiment 1 were flown to a simulated runway with a system of approach lights 900 m in length. This length falls within the full facilities system category of European aviation regulation (JAA, 2004b).

Design. The experiment used a complete block design with repeated measures on HUD use (HUD, no-HUD). The order of HUD use was counterbalanced across participants to control for potential order and carryover effects. Two blocks of visibility, the standard minimum RVR (550 m) for full facilities system and lower-than-standard RVR (400 to 500 m), were a between-subjects variable to find out whether and HUD performance advantages would also carry over into lower-than-standard RVR conditions. Each participant completed two approaches to a runway with a full system (900 m) of approach lights.

Procedure. Two approaches were manually flown using standard operating procedures. The scenarios started as a 6 nautical mile final to the runway in lower-than-standard or standard

RVR. In the HUD condition, the pilots flew the aircraft by keeping the flight-path symbol centered on the guidance cue. At 50 feet above the runway threshold, the guidance cue was automatically removed and the pilots performed the landing flare using external visual cues in conjunction with the HUD flight-path symbol. The automatic removal of the guidance cue and the HGS-4000® IMC mode were deliberately chosen to insure that the pilots could not attend solely to the HUD symbology in the with-HUD trials as well as improving the external validity (in the form of operational usability) of the study. In standard operational settings, the vast majority of the ILS systems are approved for only Category I operations and do not have the quality and integrity to support HUD modes incorporating landing-flare guidance. Approach and landing plots were printed using the aircraft's center of gravity as reference to determine the size of the touchdown footprint.

Results

The data set consists of 43 pairs of approaches. Of these 43 pairs, 18 were flow using the HUD in the first trail and 25 using the conventional head-down display (no-HUD) in the first trial; 15 were flown in standard RVR (550 m) conditions and 28 in lower-than-standard RVR conditions.

Touchdown performance. A series of full-model and reduced-model linear regressions were calculated to create separate ANOVA tables for the lateral (cross-runway) and longitudinal (down-runway) components of the touchdown footprint. The regression approach is robust to differences in sample sizes across blocks (levels of RVR) and allows for assessment of order effects and their interactions with the main factor of interest, HUD use. We adopted an alpha level of .05 for statistical significance.

Lateral touchdown performance was measured as the absolute lateral deviation from the runway centerline at touchdown. The data for the main effect of HUD use is shown in Figure 5a. There was a strong effect for HUD use, $F(1, 78) = 8.991$, $p < .004$, $\eta^2 = .12$ at an alpha of .05 indicating a power of at least .71. Landings were closer to the centerline when pilots used the HUD. The main effects for RVR and order of HUD condition were not significant. There was, however, an interaction between HUD use and the order of HUD use, $F(1, 78) = 3.235$, $p < .076$, $\eta^2 = .11$, that not achieved significance but could indicate there were asymmetric transfer effects at play. However, as shown in Figure 5b, both the group that used the HUD on their first approach (HUD 1st) and the group that used the HUD on their second approach (no HUD 1st) landed nearer the centerline when they used the HUD. Thus, there is scant evidence for asymmetric transfer effects to an extent that could have jeopardized the outcome of the experiment.

Longitudinal touchdown performance was measured as the distance from the runway threshold at touchdown. Contrary to expectations and in contrast with the results for lateral performance, the effect of HUD use was not significant. However, the between-subjects main effect of RVR, shown in Figure 6a, was significant, $F(1, 78) = 6.504$, $p < .013$, η^2 .13 indicating a power of at least .65. The mixed-effect interaction of RVR (between subjects) and HUD sequence (within subjects) was also significant, $F(1, 78) = 5.750$, $p < .019$, Figure 6b. Post-hoc analysis suggests that landings were considerably shorter in low RVR for the group that used the HUD first.

Approach success rate. Approach success rate was defined as the number of approaches, divided by the total number of approaches. A chi-square test was used to examine the independence of

landing success and HUD use. The observed data from all 125 approaches, collapsed across levels of RVR, are shown in Table 2. The test indicates that the two factors are independent, χ^2 (1) = 1.36, p > .24. Similar results were found within the low and standard manipulations of RVR.

Experiment 2

Method

The method, procedure, and design used in the second experiment were identical to those used in the first experiment with the few exceptions discussed here. The different criteria for standard RVR across facility types preclude collapsing and analyzing the data the data as a single experiment.

Participants. Twenty-eight pilots from the same major European airline volunteered to participate. None of the volunteers had participated in Experiment 1. All were qualified to fly the B737-700 aircraft and had completed their HUD training for the operator. Experience on the B737-700 varied from more than 50 to more than 1000 hours.

Simulated ground facilities. Approaches in Experiment 2 were flown to simulations of the runways at Bruxelles (ILS 25R BRU) with 600 m of approach lights or at Tromsö (ILS 01 TOS) with 420 m of approach lights. These lengths fall within the intermediate facilities category of approach lights as defined by European aviation regulation (JAA, 2004b).

Design. Two blocks of visibility, a standard minimum RVR (700 m) for intermediate facilities and a lower-than-standard minimum RVR (550-600 m), were a between-subjects variable in a complete-block repeated-measures design. Each participant completed at least two approaches

to a runway with a system of approach lights either 420 or 600 m in length. Nine of the participants completed more than two approaches each but only the first in each condition is used in the analyses of the size of the touchdown footprint.

Results

Of 28 pairs of approaches, 14 were flow using the HUD in the first trail and 14 using the conventional head-down display (no-HUD) in the first trial; 14 were flown in standard RVR (700 m) conditions and 14 in reduced RVR conditions. No differences were found across the two intermediate lengths of runway lights (420 m vs. 600 m). The analyses presented here collapse the data across this variable. The main effect for HUD use, shown in Figure 7, was the only factor in the ANOVA to achieve statistical significance. As in Experiment 1, the effect for HUD use on the lateral component of the touchdown footprint was strong, $F(1, 48) = 7.380$, $p < .009$, $\eta^2 = .06$ indicating a power of at least .54. Once again, landings were closer to the centerline when pilots used the HUD. As in Experiment 1, there was no significant effect for HUD use on the longitudinal component of the touchdown footprint. Unlike Experiment, 1 there was no effect of RVR condition. Table 3 shows the frequencies (n = 128) for the intermediate approach lights facilities (length 420 - 600 m). A chi-square test was used to examine the independence of landing success and HUD use. As in Experiment 1, the test indicates the two factors were independent, $\chi^2 (1) = 0.053$, $p > .81$. Similar results were found within the low and standard manipulations of RVR.

Discussion

There are three findings. First, HUD use per se did not influence the pilots' decision to land or go-around at the decision height. The lack of an effect of HUD suggests that the

additional information in the HUD did not distract the pilots' attention or interfere with their decision making process during the most critical portion of the approach and landing sequence. Second, HUD use significantly reduced variability in the lateral component of the touchdown footprint for all RVR conditions. Third, in contrast to its effect on the lateral component of the touchdown footprint, it appears that the HUD did not influence the size of the longitudinal footprint. The first two findings conform to our hypotheses. Here we reexamine our hypotheses about the impact of HUD use on the touchdown footprint and offer an explanation for its differential impact on the lateral and longitudinal components.

Experience engenders automaticity. For commercial pilots, hours of practice and numerous repetitions make touchdowns at standard facilities in good visibility and weather relatively routine. Nevertheless, the ubiquitous and ever-varying direction and velocity of crosswinds are likely to preclude the development of true automaticity at touchdown. Crosswinds introduce an element of uncertainty regarding drift (the lateral location of the aircraft relative to the centerline). For the pilot to detect any drift the visual ground segment needs to be long enough determine the aircraft's movement versus the ground. That means that to detect drift at all, a noticeable lateral displacement must take place and not all of this displacement can be identified before touchdown. Such exogenous uncertainty is likely to have conditioned pilots (and regulators) to tolerate a certain amount of variance in the lateral touchdown footprint.

The HUD largely eliminates uncertainty about drift. The addition of a conformal flight-path vector provides instantaneous feedback about aircraft drift and actual flight path. The additional information enables precise control of the flight path during approach and landing and reduces the variance in lateral displacement practically to nil.

The HUD mode examined provided no flare guidance and no additional information that could be used to guide the decision of when to initiate a landing flare. Control of the longitudinal component of the touchdown footprint is largely an effect of how pilots handle the aircraft's energy (operationally manifested as sink rate) in the final seconds before landing. Information used to control the aircraft's energy is provided by the optic flow from the looming runway (Lee, 1974) and, in large commercial air transport aircraft, the pilot is also aided by radio altimeter callouts that counts down from 50 feet to 0 feet (runway contact) in 10 foot decrements. The initiation of the landing flare has been shown to be a function of time-to-contact (Mulder, Pleijsant, van der Vaart, & van Wieringen, 2000). A small change in the timing of a landing flare at the nominal glide slope of 3° results in large longitudinal differences. It is easy to understand, then, why HUD use had no effect on longitudinal touchdown performance. It remains to be seen whether similar results are found for HUD modes with flare guidance, and which visual representations are actually most effective in prompting pilots to reduce sink rate at the optimal height above the runway.

Our data suggest that low RVR is likely to influence the point of touchdown. In Experiment 1, pilots consistently landed shorter in low RVR conditions than in standard RVR conditions. We attribute this observation to a change in the visual horizon. In the low RVR's used in experiment 1, the short visual segment shifts the visual horizon during the landing phase closer to the runway threshold. If the pilot uses the visual horizon as a cue to initiate the landing flare, the landing flare is likely to be initiated late. A late flare reduces the time the aircraft flies above the runway before touchdown and leads to a relatively short touchdown point. The results in our experiment match with this explanation, but further research is needed to exclude other explanations.

Low RVR interacted with HUD sequence in such a way that pilots landed shorter in low RVR if they used HUD on their first approach. The reason for this interaction is unclear but it is possible that the pilots learned and adapted to the available cues for when to initiate the flare without a HUD first and when that they could attend to HUD and outside cues when using the HUD on their second approach. In the group using the HUD first the pilots could have focused on the HUD symbology at the expense of outside cues, putting them again into an unfamiliar situation when they did not have the HUD cues available on their second approach.

The effects of HUD use on landing performance were seen in standard and lower-than-standard RVR conditions in both Experiment 1 and 2. This carries a clear and important operational implication: Based on empirical performance data, the minimum RVR for approaches using HUD could be set lower than for approaches without a HUD while still maintaining a similar or even smaller touchdown footprint.

Conclusions

Data from the experiments reported here show that HUD use improve lateral touchdown performance, likely because the conformal flight path vector makes it easier for pilots to determine, and correct for, aircraft drift. We did not find an effect of HUD use on longitudinal touchdown performance, probably because of the lack of flare guidance in the HUD modes studied here. The beneficial effects of HUD use on landing performance were seen in standard and lower-than-standard RVR conditions in both experiments, which imply that the minimum RVR for approaches using HUD could be set lower than for approaches without a HUD. Finally, we did not find any detrimental effect of HUD use on approach success rate.

These results could be critical in supporting regulatory decision making about the visibility required to commence an approach using a HUD-equipped aircraft. In fact, it is precisely this type of systematic empirical data that is often lacking as a basis for developing well-informed regulations, in Europe as well as the USA. We believe that empirical results like these should be a critical ingredient in safety regulators' rule-writing for (the application of) new technology in commercial aircraft operations. When it comes to HUD, the need for informed regulation of HUD use is already pressing, and will only increase as HUD use proliferates. The relevance of the findings here could even support the expansion of such targeted empirical studies, for example to Category II conditions, or to refine them to find, for example, the lowest acceptable RVR using HUD.

References

Fadden, S., Ververs, P. M. &Wickens, C. D. (1998). Costs and Benefits of head-up display use: A meta-analytic approach. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*,16-20, Santa Monica, CA.

Federal Aviation Authority, (2002), *Criteria for approval of Category I and Category II Weather Minima for Approach,* AC 120-29A, U.S. department of Transportation, Washington, D.C.

Fisher, E., Haines, R.F., & Price, T.A., (1980), *Cognitive Issues in Head-Up Displays*, NASA technical Paper 1711, Moffet Field, CA: NASA Ames Research Center

Flight Safety Foundation. (1991). Special Safety Report: Head-up Guidance Technology (HGST). A Powerful Tool for Accident prevention. *Flight Safety Digest*, FSF/SP-91/91, Alexandria, VA.

Gibson, J. J. (1979, 1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates:

Iavecchia, J. H., Iavechia, H. P., & Roscoe, S. N. (1988). Eye accommodation to Head-Up virtual images. *Human Factors 30,* 689-702.

ICAO. (1996). *Annex 10, Aeronautical Telecommunication, Volume I – Radio Navigation, 5th edition*. Montreal, Quebec: International Civil Aviation Organization.

ICAO. (1993). *OPS – Aircraft Operation (Doc 8168) Volume II, 4th edition*. Montreal, Quebec: International Civil Aviation Organization.

ICAO. (2001). *Air Traffic Management*, Document 4444 ATM/501, 14th Ed. Montreal, Quebec: International Civil Aviation Organization.

Joint Aviation Authorities. (2004a). *All Weather Operations JAR-AWO*. Hoofddorp, The Netherlands.

Joint Aviation Authorities. (2004b). *JAR-OPS 1 Subpart E - All Weather Operations*. Hoofddorp, The Netherlands.

Lauber, J. K., Bray, R. S., Harrison, R. L., Hemingway, J. C., & Scott, B. C. (1982). *An operational evaluation of head-up displays for civil transport aircraft.* NASA Tech. Paper 1815. Moffett Field, CA: NASA Ames Research Center.

Lee, D. (1974). A theory of visual control of breaking based on information about time to collision. *Perception 5*, 437-459.

Martin-Emerson, R. & Wickens, C. D. (1997). Superimposition, symbology, visual attention and the head-up display. *Human Factors 39*(4), 581-601.

Mulder, M., Pleijsant, J-M., van der Vaart, H., & van Wieringen, P. (2000). The effects of pictorial detail on the timing of the landing flare: Results of a visual simulation experiment. *The International Journal of Aviation Psychology, 10*(3), 291-315.

Newman, R. L. (1995). *Head-up displays: Designing the way ahead*. Aldershot: Ashgate.

Sanford, B. D., Foyle, D. C., McCann, R. S. & Jordan, K. (1993). Head-Up Displays: Effect of Information Location on the Processing of Superimposed Symbology. *Proceedings of the 7th International Symposium on Aviation Psychology*, pp. 81-87. Columbus, OH

Stuart, S. W., McAnally, K. I. &Meehan, J. W. (2003). Head-up displays and visual attention: Integrating data and theory. *Human Factors and Aerospace Safety 1*(2), 103-124

Weintraub, D. J, Haines, R. F. & Randle, R. J. (1985). Head-up display (HUD) utility, II: Runway to HUD transitions monitoring eye-focus and decision times. *Proceedings of the Human Factors Society 29th Annual Meeting*, (pp. 615-619). Santa Monica, Ca: Human Factors Society.

Wickens, C.D. & Long, J., (1995). Object versus space-based models of visual attention: Implications for the design of Head-Up displays. *Journal of Experimental Psychology: Applied 1*(3), 179-193.

Wolffsohn, J. S, Edgar, G. K., Stone, H. E., Williams, M. & Mc Brien, N. A. (1999). Does Over-Accommodation Occur When Using Aircraft Head-Up Displays? *Aviation, Space and Environmental Medicine, 70*(7), 666-673.

Author notes

Correspondence should be addressed to Örjan Goteman, Flight Operations Standards, STOOS, Scandinavian Airlines, S-195 87 Stockholm, Sweden; orjan.goteman@sas.se

The authors thank Staffan Carlsson and Zacharias Soudah for assistance in data collection, Jan-Olof Ehk for statistical advice, and the pilots who volunteered their time and skill. Rockwell Collins Flight Dynamics Company provided simulator time to conduct the experiment.

Örjan Goteman is a captain for Scandinavian Airlines flying B737. He is also responsible for specifying and implementing new aviation systems across the entire SAS aircraft fleet. He received a university medical degree from the University of Linköping in 1987 and is on schedule to receive his Ph.D. in Industrial Ergonomics from the Linköping Institute of Technology in 2005.

Kip Smith is a research professor of Industrial Ergonomics at the Linköping Institute of Technology, Sweden. He received his Ph.D. in management with an emphasis in decision science from the University of Minnesota in 1996.

Sidney Dekker is Professor of Industrial Ergonomics at the Linköping Institute of Technology, Sweden, and is Director of Studies at the Centre for Human Factors in Aviation there. He gained his Ph.D. in Cognitive Systems Engineering from The Ohio State University, USA.

Tables

Table 1. Runway visual range (RVR) standards for facilities with full and intermediate systems of approach lights.

Experiment 1 with Full Facilities (900 m)		Experiment 2 with Intermediate Facilities (420-600 m)	
Low RVR	Standard RVR	Low RVR	Standard RVR
400-500 m	550 m	450-650 m	700 m

Table 2. Contingency table of the pilots' decisions to land or go-around across conditions of HUD use at simulated facilities with a full system of runway lights (900 m), Experiment 1.

	Go around	Landing	
With HUD	9	53	62
No HUD	5	58	63
	14	111	125

Table 3. Contingency table of the pilots' decisions to land or go-around across conditions of HUD use at simulated facilities with an intermediate system of runway lights (420 or 600 m), Experiment 2.

	Go around	Landing	
With HUD	10	62	72
No HUD	7	49	56
	17	111	128

Figure captions

Figure 1.

Figure 1. The visual segment as seen from the aircraft at decision height at a given runway visual range (RVR).

Figure 2. The visual segment as seen through the left cockpit window in a B737 at decision height when flying in over a Full Approach Lights System (at least 720 m length) in (A) standard 550 m RVR and (B) lower-than-standard 450 m RVR.

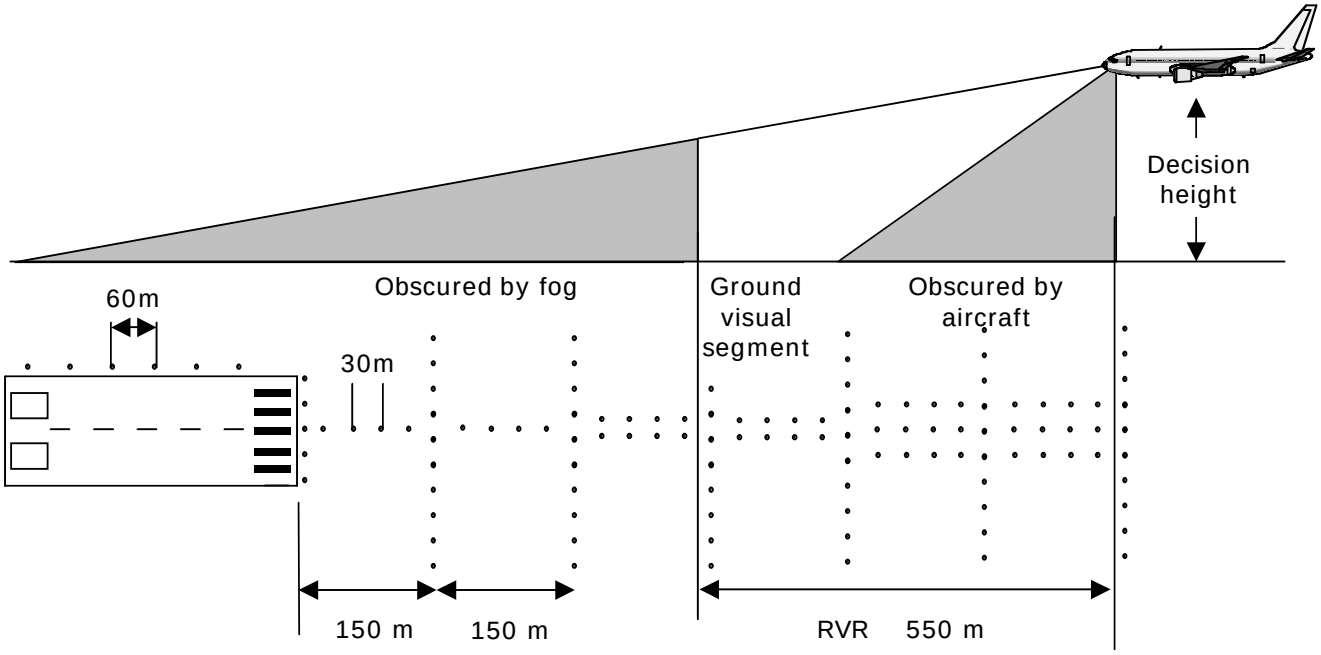
Figure 3 The visual segment seen through the left cockpit window in a B737 at decision height when flying in over an Intermediate Approach Lights facilities (420 to 719 meters length) and in (A) standard 700 m RVR and (B) lower-than-standard 550 m RVR.

Figure 4. The HUD symbology in the IMC mode used during the approaches in Experiments 1 and 2. The Flight Path symbol is a ring with slanted wings in the middle of the display. The horizontal line in the middle of the display represents the horizon. The aircraft symbol at the top of the display represents how the aircraft's longitudinal axis is pointing, in this case three degrees up and same as the actual track, no lateral drift is apparent. Aircraft speed is 130 knots and is indicated to the left in the form of a digital readout AS 130. Aircraft altitude is 900 ft barometric altitude is indicated to the right and below the horizon in the form of a digital readout as 900 B. The relative location of the flight path symbol 3 degrees below the horizon line indicates that the aircraft is descending on a 3 degree glideslope on a track of 000 degrees. Radio altitude readout is 900 ft above ground and is indicated by the digital readout of 900 just below the flight path symbol. (Courtesy of Rockwell-Collins Flight Dynamics).

Figure 5. Graphs showing (A) the main effect for HUD use and (B) the interaction of HUD use and the order of its use on the lateral size of the touchdown footprint at simulated facilities with a full system (900 m) of approach lights, Experiment 1.

Figure 6. Graphs showing (A) the main effect for RVR and (B) the interaction of RVR and trial order of HUD use on the longitudinal size of the touchdown footprint at simulated facilities with a full system (900 m) of approach lights, Experiment 1.

Figure 7. Graph showing the main effect of HUD use on the lateral size of the touchdown footprint at simulated facilities with an intermediate system (420 m or 600 m) of approach lights, Experiment 2.



HUD use Figure 1

Figure 2 (A)

Approach lights obscured by low visibility

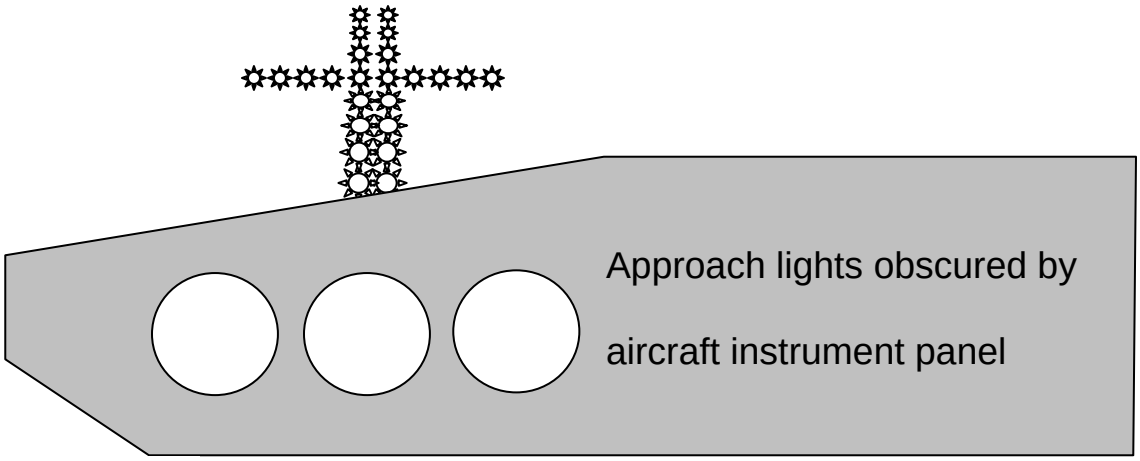
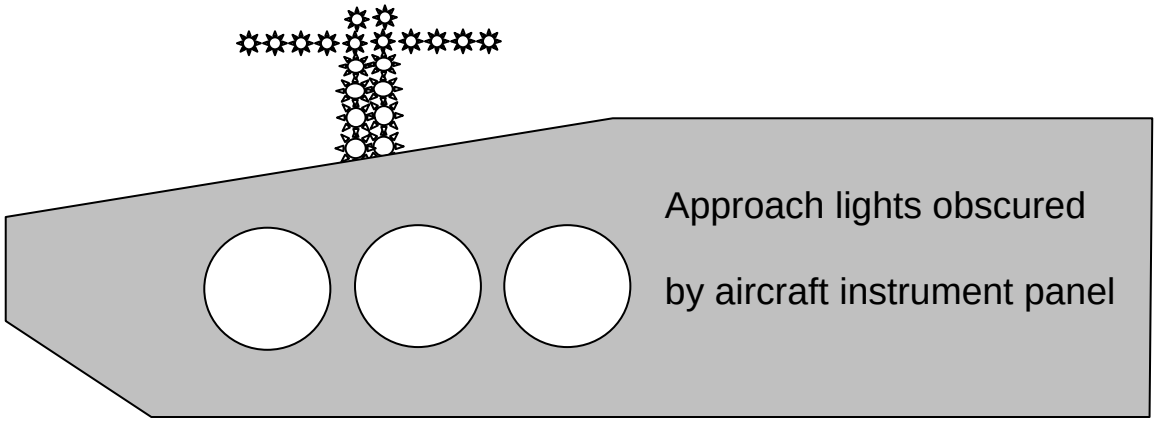


Figure 2 (B)

Approach lights obscured by low visibility



HUD use figure 2

Figure 3 (A)

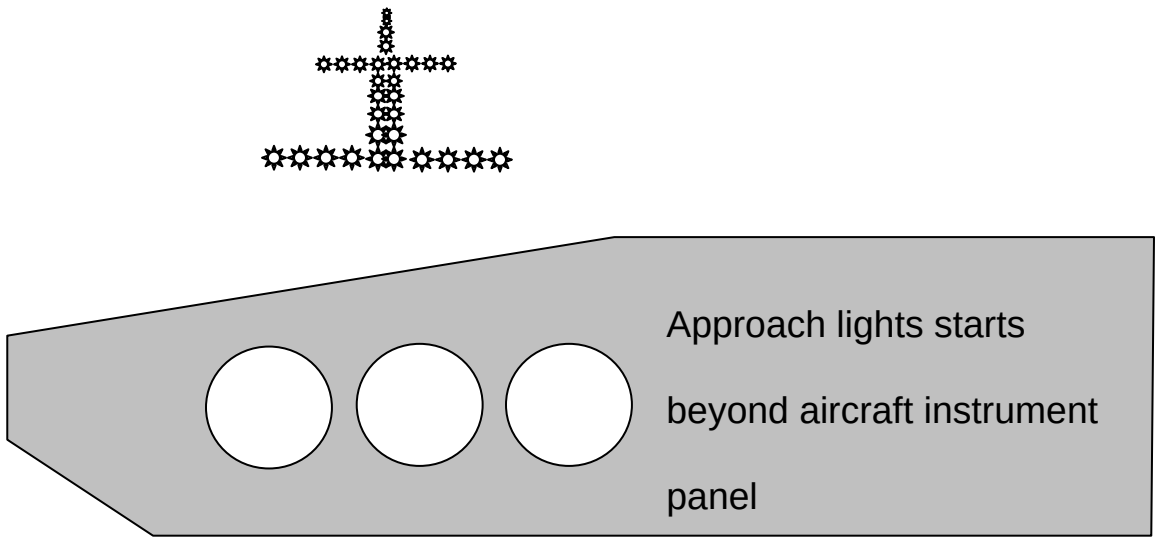
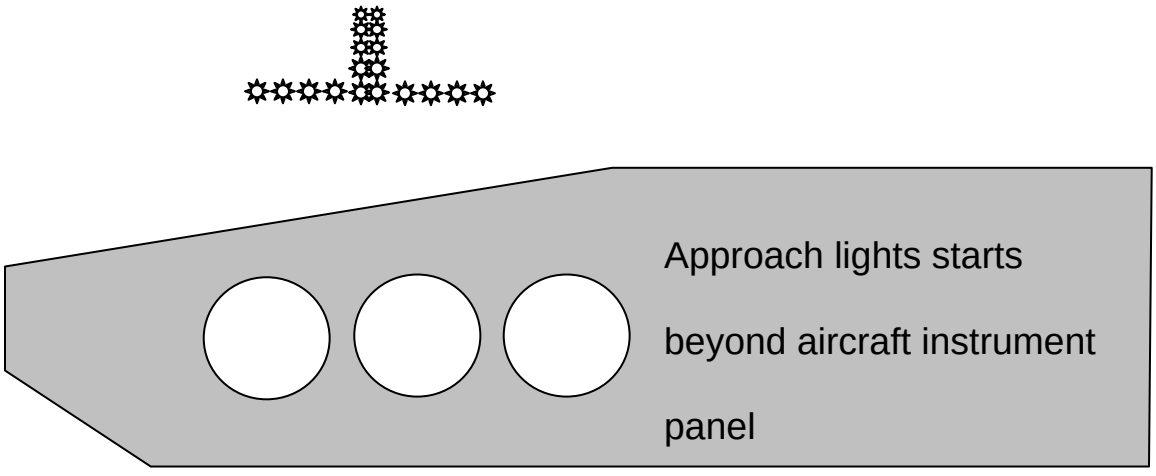
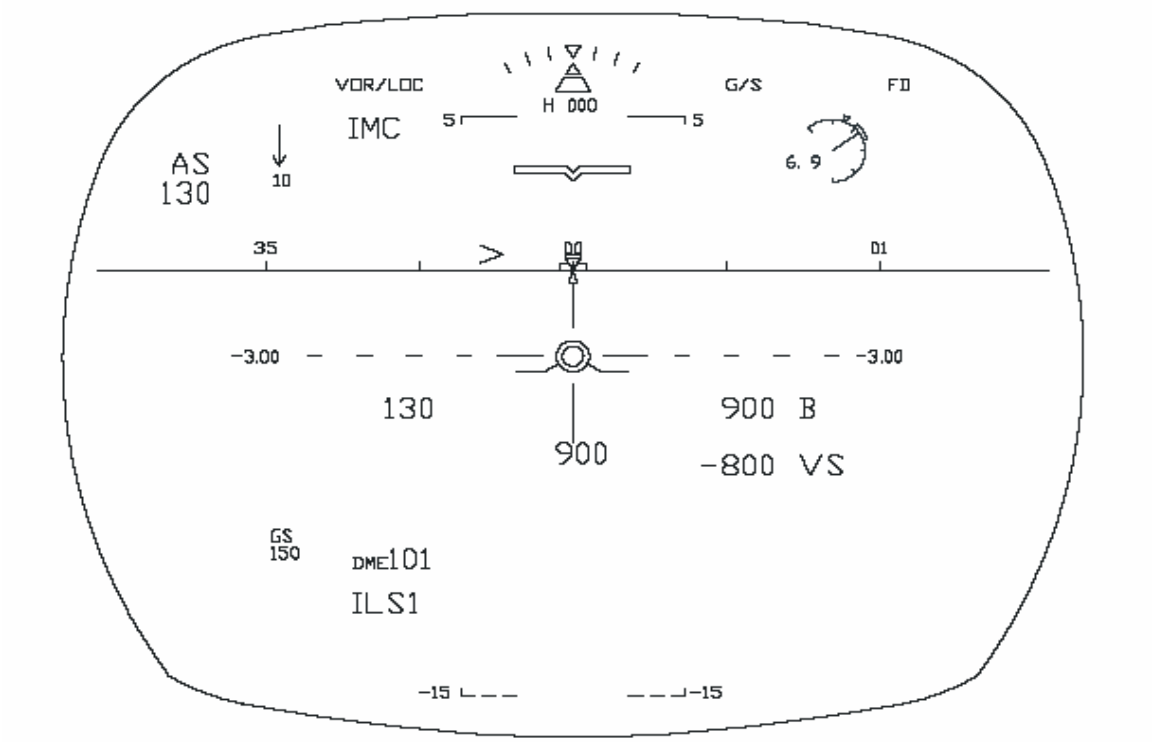


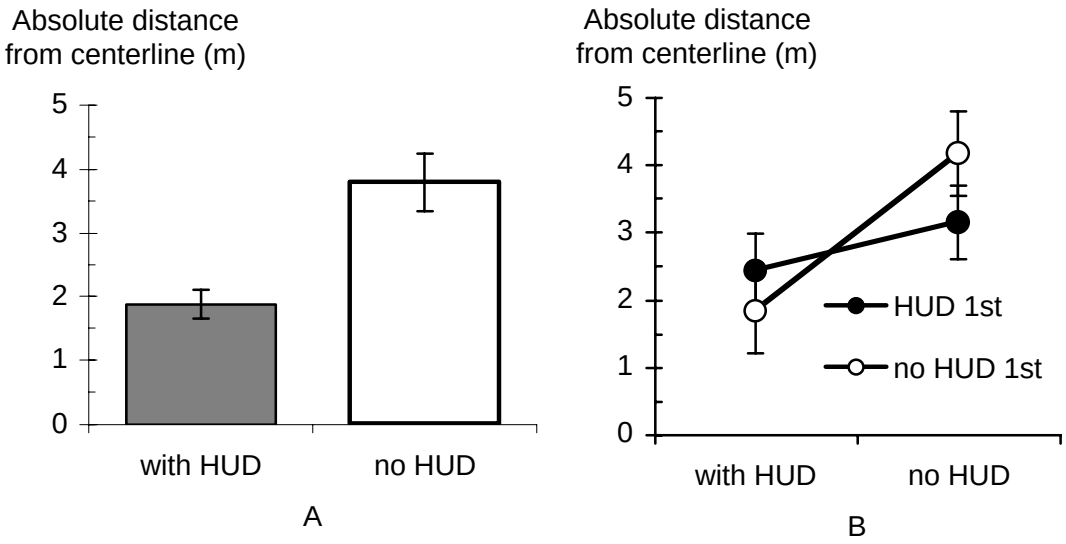
Figure 3 (A)



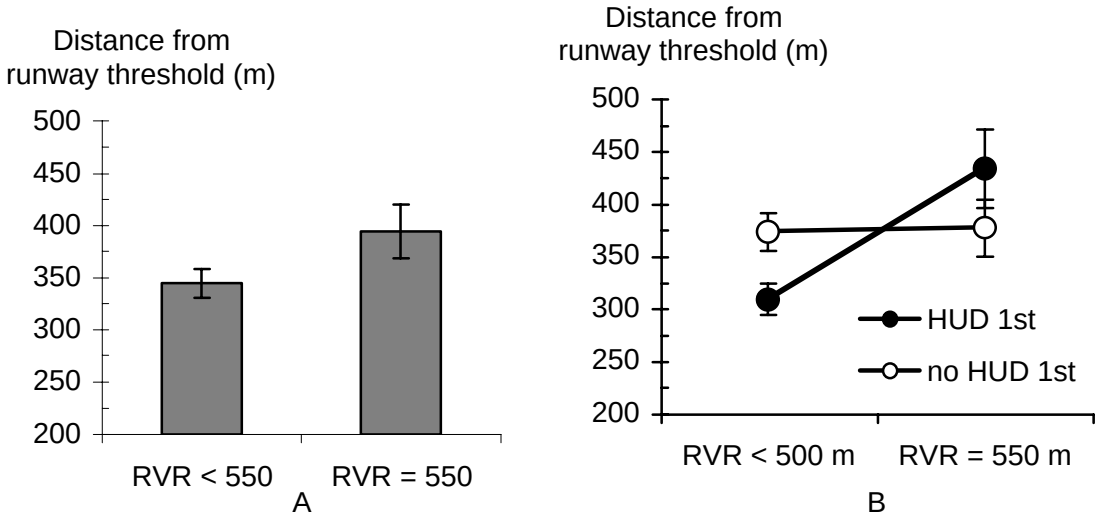
HUD use figure 3



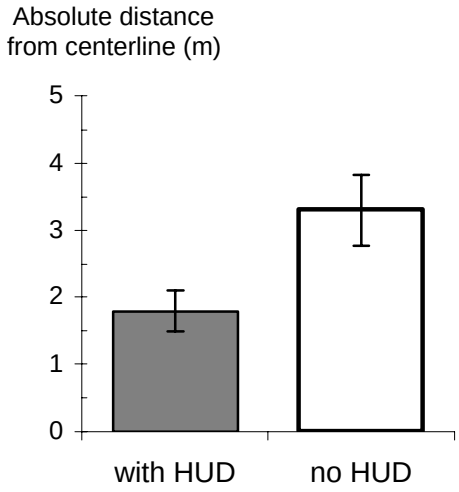
HUD use Figure 4



HUD use Figure 5



HUD use Figure 6



HUD use Figure 7

Linköping Institute of Technology Dnr IKP 214/03-23
Progress Report covering the period 1 January 2005 - 31 March 2005
MURI OPUS -- DAAD19-01-1-0621

Principal Investigator: Kip Smith, Ph.D.
Division of Industrial Ergonomics

Progress

The goal of the MURI-sponsored research at the Linköping Institute of Technology (LiU) is to examine the effects of remote command and control on soldier cognitive performance with an eye toward mitigating decrements in that performance. All effort during the first quarter of 2005 was directed at designing experiments to be conducted later in the year and at securing the collaboration of the Swedish military in those experiments.

Plans for the 2nd quarter 2005

The planned experiments will continue our investigation into the utility of augmenting cognition in remote command and control with gentle vibro-tactile stimulation. Specifically, we plan to investigate alternative explanations for why participants in our previous experiments were faster to respond to verbal orders issued by a remote leader when those orders are accompanied by a cell-phone-like rumble between the shoulder blades. We are interested in knowing whether it might be beneficial to develop vibro-tactile stimulation (simulated touch) as a medium for augmented cognition. Our application for this augmentation is non-verbal feedback between a fire team and its remote leader. The eventual outcome of this line of research might be the inclusion of ergonomic, vibro-tactile feedback devices into body armor.

In our work on simulated touch, participants (in the paintball assault lane) received two kinds of verbal orders, “move” and “fire”. Participants were faster when the verbal order was accompanied by gentle vibro-tactile stimulation between the shoulder blades. Our preferred explanation for this result is that the simulated touch stimulates trust in the remote leader.

Sharp-eyed students found a potential confound in our design: the lack of uncertainty about the nature of the next command. It is possible that the vibro-tactile stimulation removed the need for conscious processing of the verbal command. In the absence of uncertainty about the nature of the command, participants may have responded appropriately to any stimulus, regardless of modality and independent of trust. If this were the case, the advantage for vibro-tactile stimulation would presumably disappear if the participants were uncertain what the next command would be. During the 2nd quarter of 2005, we intend to test this alternative using a circular assault lane and an unpredictable sequence of three kinds of verbal orders - “left” and “right” and “fire”.

International collaboration

As part of our MURI-sponsored research program, we have been actively developing collaborative relationships with research teams in the Swedish military. Three strong bonds have been formed, all related to Sweden’s transformation to “Network-based defense” (netverksbaserad försvars, NBF).

<u>The Swedish Defense Materiel Administration (Försvarets materielsverket, FMV, www.fmv.se) and the Swedish Defense College (Försvarshögskolan, FHS, www.fhs.se)</u>

Thanks in part to our participation in the MURI project, we have been able to secure funding from the Swedish Defense Materiel Administration (FMV). The goal of the FMV project is to assess the feasibility and utility of developing a method and a tool for recognizing and

communicating constraints on distributed, collaborative decision making in dynamic command-and-control environments.

Remote command and control is the link between the MURI and FMV research agendas. While our MURI project focuses on mitigating decrements in cognition associated with remote command and control, the FMV project focuses on augmenting cognition in a distributed command-and-control center. It would be mutually advantageous for MURI-supported work at UCF (e.g., the study of hedonomics) to piggy-back upon the ongoing FMV-sponsored series of experiments at LiU.

The college and military garrison in Skövde

Thanks to the cooperation with faculty at the College in Skövde, it is likely that we will receive permission to enlist Swedish Army recruits as participants in our next series of experiments on the utility of augmenting cognition in remote command and control with gentle vibro-tactile stimulation. Our conversations with the head of the garrison in Skövde have been promising. We hope to conduct these experiments at the boot camp in Kvarn, near Linköping.

The garrison in Skövde is home to (a) a network of three immersive "virtual reality" crew-training simulators of the MBT122 (manned battle tank), and (b) the officer leadership training school for tankers. The chief of the leadership school is keen on securing our participation in a series of experiments in the networked simulators during the 3rd and 4th quarters of 2005. These experiments will likely focus on the linkage between mutual vulnerability and the formation and maintenance of trust in a remote leader. This is an exciting prospect that will build directly upon our earlier MURI-supported work conducted at Fort Riley, Kansas.

Journal articles

Reviews of two articles discussing MURI-sponsored experiments on the value of leader presence and the utility of simulated touch were received from major journals. The reviews suggest that both articles need revisions prior to resubmission. These revisions are in progress. Resubmissions will be sent during the 2nd quarter 2005.

MURI-supported and MURI-related conference presentations in 2005

Smith, K. (2005). Simulated touch may stimulate trust in a remote leader. Accepted for presentation at the 17th Annual Convention of the American Psychological Society. Los Angeles, CA. May 26-29.

Woltjer, R. & Smith, K. (2005). Constraint propagation in distributed, collaborative command and control. Accepted for presentation at the Naturalistic Decision Making Meeting. Netherlands. June 13-16.

Smith, K. & Liberg, M. (2005). Paintball 2: Empirical studies of remote-controlled tactile simulation for the stimulation of trust in remote command and control. Accepted for presentation at the 49th Annual Meeting of the Human Factors and Ergonomics Society. Orlando, FL. September 26-30.

Linköping Institute of Technology Dnr IKP 214/03-23
Progress Report covering the period 1 April 2005 – 30 June 2005
MURI OPUS -- DAAD19-01-1-0621

Principal Investigator: Kip Smith, Ph.D.
Division of Industrial Ergonomics

Executive Summary

The majority of the work conducted this quarter for the MURI-OPUS project at the Linköping Institute of Technology focused on preparing for and conducting the first in a new series of laboratory experiments on collaborative, distributed command and control. The purpose of the experiments is to investigate cultural differences in norms for decision making in collaborative, distributed command and control.

The experimental platform is the C3Fire microworld (Granlund, 2003, 2004). This memo discusses the experimental paradigm. Appendix 1 is the Instructions to Subjects used in the experiment. It contains a detailed overview of the C3Fire experimental platform.

Bugs in the experimental procedure, materials, and design were worked out in a pilot study. Thirty-two Swedish males took part in the first experiment. Mountains of useful data were acquired. Analyses are beginning.

During the quarter, work also proceded on our series of 'paintball' experiments on the effect of remote command-and-control on soldier cognitive performance. We presented a poster summarizing these experiments at the annual meeting of the American Psychological Society and had a paper accepted for presentation at the annual meeting of the Human Factors and Ergonomics Association. The HFES proceedings paper, entitled Paintball 2: A study of remote-controlled vibro-tactile stimulation for the simulation of leader presence, discusses research funded by the MURI-OPUS project and conducted in 2004. It is attached to this quarterly report. The paper will be delivered in October 2005 at the annual meeting.

Portions of the discussion of the new experimental paradigm were written by my students – Lauren Murphy (University of Central Florida), Ida Lindgren, and Rogier Woltjer – without whom little would have been accomplished. As Principal Investigator, I accept full responsibility for errors, oversights, and omissions.

If you have any questions or need additional information, please do not hesitate to contact me. I welcome your feedback.

Kip Smith, Ph.D., Gästprofessor, Principal Investigator
Division of Industrial Ergonomics, IKP/IAV
Linköping Institute of Technology

013 28 27 64 (office)

0708 48 26 22 (cell)

kipsm@ikp.liu.se

An experimental paradigm for studying the emergence of teamwork in leadership teams that are formed ad-hoc and on-site in remote command-and-control centers

Rationale

Two cornerstones of the Network-based Defense initiative (e.g., Brehmer & Sundin, 2000; Friman, 2000; Hébert, 2003) are(1) the transition to remote command and control of the dismounted infantry and (2) a renewed emphasis on multinational operations. We have embarked on a new series of experiments that are designed to investigate the effects of multiculturalism in remote command-and-control centers.

One major consequence of the transition to multinational operations and remote command-and-control is that soldiers in the field may no longer take their orders from an officer whom they can see. Instead, orders will be given by leaders located in a safe position away from the battlefield. In our 'paintball' line of research we have been and will continue to focus on the effects of this transition on soldier performance. In the new series of experiments, we are focusing on teamwork in the command-and-control center.

The multinational leadership teams are likely to be formed ad-hoc and on-site. Team members may or may not know each other. Because there is often little time for team-building, the team will get to know each other as they work. As they get to know each other, their way of working together is likely to evolve. Their task is to collaborate and to make decisions to coordinate strategic and tactical operations.

The goal of our new experimental paradigm is to create a laboratory setting and gather data that make it possible to document and understand the emergence of teamwork in multicultural leadership teams that are formed ad-hoc and on-site. This work is directly relevant to the goal of the MURI-OPUS program: to document and understand sources of decrements in cognitive performance associated with the transition to Network-based Defense and remote command and control.

To address this goal, we have designed an experimental paradigm that:

(1) elicits and captures spontaneous but collaborative decision making that coordinates strategic and tactical operations in response to a simulated emergency,

(2) emulates the potentially ad-hoc nature of leadership teams in a remote command-and-control center and to capture the actions, words, expectations, and norms for decision making that guide the development of teamwork, and

(3) gathers self-report information about participants' preferences for team organization, roles and priorities. The self-report also aims at finding personality traits which correlate with cultural differences documented by others (e.g., Hofstede, Markus, Triandis).

The structure of the experiment was a cycle containing three parts – team play using the C3Fire mircroworld (Granlund, 2004), team discussion of their play, and individual responses to questionnaires. The three parts of the cycle correspond to the three goals enumerated above. Over the course of two days, team of four randomly assigned participants cycled between parts at total of eight times.

In the first part of each cycle, teams of four worked together across a computer network to coordinate emergency services operations posed by the C3Fire microworld. C3Fire recorded the actions they took and their email communication.

In the second part of each cycle, the team sat around a table, watched a replay of their C3Fire session, and engaged in an open-ended conversation about their play. During these conversations, the team developed an organizational structure, allocated responsibilities, and debated alternative strategies for dealing with the emergencies posed by C3Fire. This conversation was recorded with a videocamera and audio equipment for subsequent qualitative and quantitative analyses.

In the third part, members of the team individually filled out one or more of a series of questionnaires. The questionnaires were drawn from literature on social, clinical, or cultural psychology.

The discussion that follows reflects this tri-partite structure.

Pilot study

All three parts of the method were initially tested in pilot study. Four Swedish males drawn from a convenience sample volunteered to participate in exchange for movie tickets. They engaged in eight cycles consisting of (1) a session with C3Fire immediately followed by (2) team discussion and (3) questionnaires.

Play with C3Fire in the pilot study led us revise aspects of the experimental scenarios to make them progressively more challenging. The use of email communication in the pilot study was spare. Accordingly, we modified our plans to put a greater emphasis on the team discussion sessions. We tested using C3Fire's replay function during these sessions in an attempt to encourage fruitful discussion about organization, strategies, and tactics. Participants found it easier to talk about the C3Fire session if they were able to watch it at the same time. The replay was a success and became a standard part of the experimental procedure.

After modifying the C3Fire scenarios and upgrading some of the equipment used to record the audio and video data, we began to solicit volunteers to participate the first experiment.

Participants

Thirty-two Swedish males (mean age 24.56, range 20 to 37) participated in the study for monetary compensation.

Male participants were chosen for several reasons. First of all, it may become difficult and inappropriate to mix men and women in some of the cultures we are interested in observing. The choice was therefore between a homogeneous female or male group. Since we will have both male and female experiment leaders, we asked the member of the reference group from Migrationsverket about the (in-) appropriateness of having a male group with a female experimenter versus a female group with a male experimenter. She told us that in strict Islamic cultures, the first option was preferable. Strictly Islamic men who have immigrated to Sweden have been in contact with many Swedish women (Migrationsverket, when learning Swedish, Arbetsförmedlingen, etc.) and are used to being around women with some authority. Their

female counterparts have the same kind of experience with Swedish women relatively little experiment with Swedish men. Accordingly, strictly Islamic women will probably not be comfortable in the presence of male experimenters.

A wish to avoid the gender issue was an additional reason to keep the groups homogeneous. The ongoing discussion of the social impact on the creation of gender versus biological implications of the sexes is lively and intense. We want to eliminate the risk of ending up in a discussion about differences between men and women as a confound in our experiment. Mixing the sexes would add yet another dimension to the analysis and might complicate the differentiation between cultural and individual differences.

The32 Swedish males were all over 20 years of age and indicated a willingness to participate in two four-hour sessions. They were promised 500 sek in compensation if they completed the full eight hours of experimentation.

The attempt to attract participants through rewarding them with tokens of gratitude and gift certificates, as discussed during the reference group meeting, turned out to be impossible. There are tax laws which make such gifts taxable. If a participant were to receive a gift, he would incur a tax obligation. Furthermore, we had to raise the compensation to 500 sek to entice participants to take part of our study. Note: this 500 sek is also taxable but participants appear to find it less aversive to pay tax on cash than on gifts or coupons.

The 32 participants reported to the laboratory in four groups of eight. In the laboratory, the eight were randomly and anonymously assigned to two leadership teams of four. The purpose of the random and anonymous assignment to teams was to minimize reputation effects. In what follows, the word ‘team’ signifies the four participants in a simulated remote command-and-control center and is the basic unit of analysis in the experiment. The word ‘group’ is reserved for all eight participants when the two teams are brought together.

The two teams of four worked in parallel but were in no way linked. This arrangement made it possible to gather data on two teams (two units of analysis) simultaneously. It also provided the opportunity to periodically reassign participants to teams, again to minimize reputation effects.

Apparatus

Part 1 – C3Fire sessions

The C3Fire sessions were conducted over a server-client network of computers in a laboratory on the LiU campus. Each of the eight participants sat a separate client computer and was linked to his teammates by the email system provided by the C3Fire software on the server computer for his team. The two teams of four were connected to different server computer. The two servers independent ran the same C3Fire scenario concurrently.

C3Fire (Granlund, 2002) is a fire-fighting microworld in which participants play the roles of fire-fighting unit chiefs and commanders. Their task is to collaborate to fights fires in an experimentally-controlled setting under observation of an experiment manager. C3Fire captures many of the characteristics of complex environments: It implements high degrees of connectivity, complexity, uncertainty, time pressure, and polytely (multiple, concurrent goals)

(Brehmer, 2005; Brehmer & Dörner, 1993; Funke, 1993, 2001; Grey, 2003; Rigas, Carling, & Brehmer, 2004).

The C3Fire interface and instructions on how to interact with C3Fire are discussed in depth in the Instructions to Subjects, Appendix 2 of this report. An example of the interface is shown the figure on the last page of Appendix 2. The interface contains a map, the email facility, and information about fire-fighting equipment. The map is divided into a grid of squares. All members of a team saw the same information on the interface.

The emergency situation is a forest fire. The speed of burning and spreading of the fire are functions of vegetation, terrain, the presence of buildings, and wind direction and speed and are pre-set by the experiment manager. Participants can extinguish the fire by placing fire trucks on top of the squares (of the map grid) that are on fire. The trucks are constrained by limits on their driving, deployment, and extinguishing speeds, as set by the experimenter. The participants tell the trucks where to go.

In previously published studies of network-based command and control using C3Fire, the fire trucks were the only trucks under the participants' control (Artman, 1999; Granlund, 2002, 2003; Johansson, 2003; Johansson, Persson, Granlund, & Mattsson, 2003; Svenmarck, 1998). This restriction facilitated analysis of the effects of different configurations for command and control interaction on the decision making process. Our focus is less on the effects of organizational structure than on the interplay of a distributed group of decision makers. To this end, we extended C3Fire to include additional command functions and four new classes of fire-fighting units.

Two of the new classes of units are water and refueling trucks. Water trucks supply fire trucks with water, and obtain water from water stations. Refueling trucks supply both fire trucks and water trucks with fuel and obtain fuel from fuel stations. The third and fourth new classes of units are water and fuel stations. Their limited resources must be allocated across trucks and cannot be moved once established. They can, however, be consumed by the fire. The addition of water and refueling trucks and stations makes the logistics of the simulation emulate some of the complications experienced by members of a remote command-and-control center.

Interdependencies among decision makers arise whenever different classes of fire-fighting units are assigned to different participants in the simulated remote command-and-control center. For example if fire trucks need water and fuel, water trucks and refueling trucks constrain the actions of fire trucks. If different people have control over these different units, their actions are mutually constraining. This provides ample opportunity for conflicts to arise.

Every event in the simulation generates time-stamped data. Truck movements, actions of decision makers, and communications between decision makers form a vast amount of data that is generated and stored. The current project aims to analyze these data and extract the norms for decision making that decision makers implicitly or explicitly use to coordinate their emergency-services activities. The data available for such analysis include:

(a) The allocation of responsibilities across participants. For example, in one allocation, each participant might be responsible for a clearly-specified set of fire-fighting units defined by type or location or both. In counterexample, the participants may choose to share responsibility for

some or all units. There is a continue of allocations between these extremes. Consistent use of a preferred allocation is evidence for a tacit norm for decision making.

(b) The actions the decision makers take. Patterns of actions reveal strategies for coping with the emergency. Some strategies require more cooperation than others. The relative amounts of collective action and of individual action are expected to correlate with well-documented metrics of cultural difference being assessed by a three different questionnaires.

(c) The content and tone of the participants' emails and conversations. The same information can be delivered as an order or as instructions or as a request. Both the content and temper of these communications are likely to reveal a lot about tacit norms for collaboration.

Part 2 - Discussion sessions

The discussion sessions took place in the laboratory. The team members gathered around the server monitor where the replay was shown. A couple of warm up questions were posed by the experimenter and the participants were allowed to discuss whatever came up. The discussion was captured with a digital video camera (Samsung, VP-D903i) and an mp3-player (Asono, Opus, 512MB). The camera and mp3-player was connected with a ZZ mixer table. The mixer was in turn connected to two external microphones. This setup guarantees that the sound recorded on the camera and the mp3-player is the same. The audio files are in mp3 format. The videotaped material will be transferred to wmv-format (Windows Media Audio/Video File). All files are possible to process and manipulate in a computer using commercial software.

The discussions (and the C3Fire email correspondence) will be transcribed and, when necessary, translated to Swedish. The transcripts are the basic data for a wide range of qualitative analyses. The primary tools for these analyses will be traditional protocol analysis (Ericsson & Simon, 198x) and content analysis (Neuendorf, 2002). Topics to be assessed include, but are not limited to:

1) organizational structure as revealed by (a) patterns of communication and (b) semantic content.
2) allocation of roles and responsibilities,
3) conflict avoidance as revealed by (a) acquiesce to tacit authority and (b) assertiveness,
4) tolerance for uncertainty as revealed by the sharpness of the definition of roles and division of responsibilities,
5) time horizon as revealed by a tactical, responsive focus or by a strategic focus, and
6) power distance as revealed by organizational structure and role allocation.

The unit of analysis will be both the team and the individual. The strength of the correlation between the analyses at the level of the individual participant and that participant's responses to questionnaires (below) will for the basis for assessing individual differences in our sample. Differences across teams from the same culture will serve as a measure of intra-cultural variability. Difference across teams from different cultures will highlight contrasting norms for collaboration and decision making. The fundamental premise of the research is that these contrasts have the potential to raise barriers to collaborative command-and-control operations.

Part 3 – Questionnaires

The team of participants that is not engaged in a discussion session fills out a questionnaire or two. There are a total of 7 questionnaires.

Demographic questionnaire

The demographic questionnaire is a 16-item self-report questionnaire with both open-ended and forced-choice questions which assesses the participants' (1) personal, academic and work related background, (2) international experiences, (3) estimated language skills, (4) experiences of military, emergency service or police work, and (5) experiences of using computers, especially word processing and chat programs.

NEO Five Factor Inventory (NEO-FFI).

The NEO-FFI is designed to measure the 'Big Five', five domains of adult personality: extraversion, agreeableness, conscientiousness, emotional stability, and intellect/openness. It will give us valuable insights in the participants' personalities and will facilitate differentiation between individual and cultural differences. It is one of our main tools for fending off objections that our data may be dominated by individual rather than cultural differences.

The NEO-FFI is a 60-item personality inventory (Costa & McCrae, 1992). We have obtained a license to administer the inventory and are paying its copyright holder a per-participant fee. The NEO-FFI has adequate internal consistency, construct, and discriminative validity across diverse samples (Ball, Rounsaville, Tennen, Kranzler, 2001; Costa & McCrae, 1992).

Conflict Avoidance (CA).

In some cultures, the wishes and welfare of the collective is more important than the wishes and welfare of the individual (e.g., Markus & Kitayama, 1991). Collectivist cultures seek harmony. In individualistic cultures, the focus is the welfare of the individual. There is likely to be more conflict between individuals in an individualistic culture than a collectivist culture. The CA will allow us to place participants along the individualism-collectivism continuum.

The CA is a 23-item self-report scale that assesses a person's disposition to react to conflict. The scale is comprised of items from the NEO Pi-R, the ICAPS, the ROAD, and an additional Conflict Avoidance scale taken from (Tjosvold, 1985; Barker, Tjosvold, & Andrews, 1988).

Parts of the CA scale can also be related to the cultural dimensions known as 'Power Distance' and 'Masculinity' (Hofstede, 2001). For example, Swedish people are known for their wish for consensus in decision making (low power distance and exceptionally low masculinity). Consensus is thought to make more individuals happy with the decision. However, the process can be time consuming. In cultures in which a person in authority makes most of the decisions and his subordinates follow (high power distance and high masculinity), the decision making process can be swift. In such cultures the Swedish way of seeking consensus can be experienced a frustrating barrier to collaboration and decision making. The CA questionnaire is designed to capture differences of this kind.

Tolerance for Uncertainty (TU).

People react differently to situations in which they feel unfamiliar or without complete information. The TU questionnaire is designed to capture a person's level of acceptance for the unknown, i.e. to see if they become stressed in situations where they do not understand everything and they are unable to predict what will happen next. It is another of our tools for fending off objections that our data may be dominated by individual rather than cultural differences.

The TU is a 24-item self-report scale that assesses the degree of comfort in making decisions with incomplete information and in unfamiliar situations. The scale is comprised of items adopted from the Need for Cognitive Structure scale (NCS; Bar-Tal, 1994), the Revised NEO Personality Inventory (NEO-PI-R, Costa & McCrae, 1992) and the Uncertainty Response Scale (URS; Greco & Roger, 2001). The URS is comprised of three factors, Emotional Uncertainty (EU), Desire for Change (DC), and Cognitive Uncertainty (CU). All three sources have satisfactory internal consistency and test-retest reliability across diverse samples (Ball et al., 2001; Bar-Tal, 1994; Greco & Roger, 2001).

Time Horizon (TH).

People from different cultures can differ in how far ahead they plan and how much control they feel they have over their own lives. On the one extreme is the individual who believes he has full control over his life and plans far ahead. On the other end are the individual who believes that he is in the hands of fate/God/Allah/etc., who thinks that it is impossible to know what tomorrow brings, and who sees no point in planning ahead. Imagine the barriers to communication and planning when these two are forced to work together. We expect their norms for planning and decision making will collide dramatically. The TH questionnaire is a tool for assessing differences of this kind.

The TH is a 19-item self-report scale that assesses how far people plan in advance, the time frame of their goals, and how far ahead they look to justify their actions. The scale is comprised of items from the Uncertainty Response Scale (URS; Greco & Roger, 2001), the PFI, the 'Time Orientation' dimension of the Value Orientation Method Survey (VOM; Kluckhohn & Strodtbeck, 1961; Kluckhohn Center, 1995), and the Ability to Achieve Cognitive Structure Scale (AACS; Bar-Tal, 1994). The AACS has both satisfactory internal consistency and test-retest reliability (Bar-Tal, 1994).

Flow State Scale (FSS).

A 36-item Flow State Scale, developed by Jackson and Marsh (1996), was used to assess degree of flow experienced while engaged in a particular activity. Flow is the subjective experience of connectedness with events and the environment (Csikszentmihalyi, 1990, 1997). The FSS has received initial psychometric support including adequate construct validity (Jackson & Marsh, 1996).

Chronometric analysis of changes in responses to the FSS will reveal patterns in the participants' involvement with the experimental task and, by inference, the evolution of cohesion in the command-and-control team. We intend to perform correlation analysis of the FSS and behavioral measures. Teams that acquire the experience of flow are expected to work together seamlessly while teams that resist flow are expected to have trouble communicating.

State-Trait Anxiety Inventory (STAI).

The STAI is a 40-item self-report scale assessing current (state) and personality (trait) anxiety (Spielberger, 1983). Only two dimensions of the STAI are being used for the study, the 14-item 'Opinion of the Task' dimension and the 20-item 'Dealing with Problems' dimension. The STAI has high internal consistency and adequate construct and discriminative validity across diverse samples (Novy, Nelson, & Goodwin, 1993).

The STAI will give us insight into a participant's relative level of involvement in the experimental task and preference for dealing with potentially stressful situations. Consistent patterns of preference within a cultural group may be a diagnostic of differential cultural norms for decision making.

Design

Anonymity

One of the salient social characteristics of the Network-based Defense initiative is that the multinational leadership teams in a command-and-control center may not know each other when they arrive on site. They get to know each other as they work. There is no straightforward way to capture this emergence from anonymity in an experimental setting in which participants may indeed know each other. We crossed this hurdle by bringing eight participants into the laboratory at once and splitting them, randomly and anonymously, into two groups of four. Initially, no one on a team knew who the other three team members were. As they day advanced, they got to know each other and had numerous opportunities to interact as a team.

Specifically, team membership is unknown during the first two C3Fire scenarios. The team must attack the emergency posed by C3Fire at once and without consultation. They learn about the team and find ways of fighting the fire during the first two scenarios. After the second scenario, they are introduced to each other and engage in watching and discussing a replay of the last played scenario. During this conversation, they "get to know" each other. This interaction provides us with the opportunity to observe how the team's behavior may change as a function of their knowledge about each other.

Scenarios

Table 1 shows the systematic manipulation of C3Fire variables across the eight experimental sessions. Each C3Fire scenario lasted until the participants completely suppressed the fire or 20 minutes elapsed, whichever came first.

Participants encountered scenarios designated A through D were on the first day of experimentation. They encountered scenarios E through H on the second day. Two different maps with differing configurations of forests and houses, etc., form the foundation for the eight scenarios. Four scenarios use map number 1 and four use map number 2. The maps are systematically rotated to make them appear different. The third column of the table indicates the angle of rotation in degrees. 'Initial fire size' indicates the size of the fire, in squares, at the beginning of the scenario. The term 'Relative challenge' refers to an ordinal scale of assumed

challenge with following relationships: level 1 < level 2, level 1 < level 3, level 2 < level 4, and level 3 < level 4, where < means 'is less challenging than.' For example, scenarios A and B are assigned the same ordinal level of challenge because the only difference between them is the map. Scenario A (level 1) is assumed to be less challenging than scenario C (level 2) because it has a smaller initial fire size on the same map. Similarly, a smaller fire size is the basis for assuming scenario E (level 3) is less challenging than scenario G (level 4). Four additional variables are manipulated in the scenarios for the second day to increase the challenge and to reduce the likelihood that participants will be able to predict how the emergency will develop. The additional manipulations make it likely that the scenarios in the second day will be more challenging than their counterparts during the first day, e.g., scenario E (level 3) should be more challenging than scenario A (level 1). However, we cannot know *a priori* the rank order of levels 2 and 3. Casual observation and an initial review of the data collected in the first set of experiments suggest that we succeeded in making the scenarios more challenging the second day.

Table 1. Dependent variables and their manipulation in the C3Fire scenarios.

Scenario Designation	Map Number	Map Rotation Angle	Initial Fire Size	Relative Challenge Level	Additional Manipulation
Day 1 Scenarios					
A	M1	0	2x2	1	
B	M2	0	2x2	1	
C	M1	90	3x3	2	
D	M2	90	3x3	2	
Day 2 Scenarios					
E	M1	180	2x2	3	faster fire spread rate
F	M2	180	2x2	3	slower vehicle speeds
G	M1	270	3x3	4	slower fire suppression
H	M2	270	3x3	4	smaller fuel tank size

Procedure, Day 1

Table 2 outlines the experimental procedure for the first day of experimentation. The day began with three C3Fire training sessions. Each participant then worked as a member of a team of four to manage the emergencies posed by C3Fire scenarios A through D, Table 1.

To capture the ad-hoc nature and relative anonymity of OSOCC formation, the participants were assigned numerical identities (i.e., 1, 2, … , 8) and player names (A, B, C and D) upon arrival. The numerical identities were meant to ease the handling of questionnaires and registration of data. The player names were assigned automatically by C3Fire and were used keep track of which participant was which player during the scenarios. Participants were instructed to keep their identities and player names secret.

Table 2. Outline of the procedure for day 1. All discussions were video- and audio-recorded. The order of questionnaire presentation was counterbalanced.

Introduction

The participants are seated. Informed consent forms are handed out, explained, and signed.
The participants are told not to talk about the game or their identities with each other until the experimenters have given their approval.
Demographic questionnaire

Instructions and training

Reading the instructions.
C3Fire training scenarios 1 and 2: Individual training
The 8 participants are split into 2 teams of 4 randomly and anonymously and work as teams.
C3Fire training scenario 3.

Cycle 1

The group is randomly and anonymously divided into two new teams and is told so.
C3Fire scenario A.
FSS and STAI questionnaires
Structured group discussion and coffee

Cycle 2

The group is randomly and anonymously divided into two new teams and is told so.
C3Fire scenario B.
FSS and STAI questionnaires
Team replay and discussion / NEO-FFI, TU, TH or CA questionnaires

Cycle 3

The teams are in the same as in scenario 2 and are told so.
C3Fire scenario C.
FSS and STAI questionnaires
Team replay and discussion / NEO-FFI, TU, TH or CA questionnaires

Cycle 4

The teams are in the same as in scenarios 2 and 3 and are told so.
C3Fire scenario D.
FSS and STAI questionnaires
Team replay and discussion / NEO-FFI, TU, TH or CA questionnaires

Thank you and good bye

Reminders to return for the second session

Instructions

Participants are asked to sit at their computer stations, to sign the informed consent form and fill in the demographic questionnaire. The participants can chose computer stations as they wish. To be able to keep track of which individual was which player during the scenarios, each participant is given a list of which player name to chose for the different scenarios. They are not allowed to tell their player names to other participants. Participants are then asked to look at their computers and study the C3Fire interface and to read silently the instruction sheet as the experimenter read the instructions aloud (see Appendix 2 for the Instruction to Subjects form).

Practice sessions

Participants then begin and complete the three practice sessions using C3Fire. The training sessions are constructed to help the participants get acquainted with the task and the C3Fire world. The first two training sessions are played individually. Each lasts for approximately two minutes. The participants are then randomly and anonymously divided into two teams and played a training session together for approximately five minutes to get a feel for how to work together and communicate using the C3Fire email system.

First cycle

C3Fire

After the practice sessions the participants were randomly assigned to two teams. At this point, the identity of team members was anonymous. The teams managed the emergency posed by C3Fire scenario A.

Questionnaires and discussion

After the scenario, all participants individually completed the FSS and the STAI; the order in which they received the questionnaires was counterbalanced. Upon completing these questionnaires, participants were asked to help themselves to coffee and cookies. The entire group of eight returned to their seats at their computers and engaged in a 10-minute long structured group discussion on C3Fire and general strategies. The experimenters kick-started the conversation by asking:

1 How did it go?
2 Did anything unexpected happen? What?
3 Why was it unexpected?
4 If you had additional information would it have been easier for you?
5 Did you use the email system? Did you have any trouble with it?
6 What information does the team need to share?

They were allowed to talk about the specific scenarios and their identities but this information was not elicited by the experimenters. The conversation was supervised and structured by the experimenters and was video- and audio recorded.

Second cycle

C3Fire

To retain anonymity, the participants were randomly assigned to two new teams. The purpose of this manipulation was to preclude domineering behavior by any one individual and to emulate the ad-hoc nature by which multinational command-and-control teams will likely be formed. The new teams managed the emergency posed by C3Fire scenario B.

Questionnaires and discussion

After the scenario, participants individually completed the FSS and the STAI; the order in which they received the questionnaires was counterbalanced. Team 1 was then gathered to watch a

replay of the scenario they just played and to discuss the replay. The experimenters started the discussion using the same questions as those used during the previous structured group conversation. The team was allowed to talk about the replay however they wanted. The conversation was video- and audio recorded.

Meanwhile, members of Team 2 were instructed to complete one of the NEO-FFI, TU, CA, and TH questionnaires. The order in which they received the questionnaires was counterbalanced.

During this time, coffee was made available. When the teams finished the discussion / questionnaires, they switched roles. Team 2 watched and discussed their replay and Team 1 completed one of the NEO-FFI, TU, CA, and TH questionnaires a counterbalanced order.

Third cycle

At this point, team membership was no longer anonymous. Like members of a multinational team, our teams were getting to know each other as they worked.

The teams then worked together to manage the emergency posed by C3Fire scenario C. Following the game, they engaged in reply/discussion sessions and completed questionnaires. The procedure was identical to that for cycle 2 except that Team 2 watched the replay first. When Team 2 finished the discussion and Team 1 finished the questionnaires, they switched roles.

Fourth cycle

The fourth cycle followed the same procedure as the third except that Team 1 watched the replay first.

Thank you and goodbye

At the end of the fourth cycle, participants were thanked and reminded to return the next day for another session if they wanted to receive they remuneration. The first day of experimentation lasted approximately four hours.

Procedure, Day 2

Table 3 outlines the experimental procedure for the second day of experimentation. Each participant worked as a member of a team of four to manage the emergencies posed by C3Fire scenarios E through H, Table 1. The procedure for the second day was essentially similar to that for the first day. When the participants arrived, the group of eight was again divided randomly and anonymously into two teams of four.

Table 3. Outline of the procedure for day 2. All discussions were video- and audio-recorded. The order of questionnaire presentation was counterbalanced.

Introduction

Short introduction..

Cycle 5

The group is divided into two new teams and are told about so. They are told that the fire spreads faster in this scenario.

C3Fire scenario E.
FSS and STAI questionnaires
Structured group discussion and coffee

Cycle 6

The group is divided into two new teams and are told about so. They are told that the trucks would move more slowly in this scenario.

C3Fire scenario F.
FSS and STAI questionnaires
Replay / Coffee

Cycle 7

The groups are in the same teams as in scenario 6 and are told so. They are told that it takes longer to suppress a fire in this scenario.

C3Fire scenario G.
FSS and STAI questionnaires
Replay

Cycle 8

The groups are in the same teams as in scenario 6 and 7 and are told so. They are told that the fire and water trucks have smaller fuel tanks in this scenario.

C3Fire scenario H.
FSS and STAI questionnaires
Replay

Thank you and good bye

Payment
Open discussion about the game and their experiences of C3Fire and the experiment.

Instructions and review

Participants were given no instructions or training but were told that specific characteristics of the fire or of the vehicles would differ from scenario to scenario. They were not told that these manipulations, shown in Table 1, were designed to increase the challenge. They were encouraged to listen attentively to the instructions preceding each scenario. The participants were reminded not to talk about the game or their identities with each other until the experimenters had given their permission to do so.

Fifth cycle

C3Fire

Participants were randomly assigned into two teams. At this point, the identity of team members was anonymous. The teams managed the emergency posed by C3Fire scenario E.

Questionnaires and discussion

After the scenario, all participants individually completed the FSS and the STAI; the order in which they received the questionnaires was counterbalanced. Upon completing these questionnaires, participants are asked to help themselves to coffee and cookies. The entire group of eight returned to their seats at their computers and engaged in a 10-minute long structured group discussion on C3Fire and general strategies. As they did during Day 1, the experimenters kick-started the conversation by asking a set list of questions. The participants were allowed to talk about the specific scenarios and their identities but this information was not elicited by the experimenters. The conversation was supervised and structured by the experimenters and was video- and audio recorded.

Sixth cycle

C3Fire

To retain anonymity, the participants were randomly assigned to two new teams. The new teams managed the emergency posed by C3Fire scenario F.

Questionnaires and discussion

After the scenario, participants individually completed the FSS and the STAI; the order in which they receive the questionnaires was counterbalanced. Team 1 was then gathered to watch a replay of the scenario they just played and to discuss the replay. The experimenters started the discussion using the same questions as those used during the previous structured group conversation. The team was allowed to talk about the replay however they wanted. The conversation was video- and audio recorded.

Meanwhile, members of Team 2 were instructed to wait in an adjacent room where they relaxed and drank coffee after being reminded not to talk about the game. The teams traded places when Team 1 finished their discussion.

Seventh cycle

At this point, team membership was no longer anonymous. The teams then worked together to manage the emergency posed by C3Fire scenario G. Replay and waiting sessions (like those in cycle 6) were repeated with the exception that Team 2 watched the replay first.

Eight cycle

The procedure for cycle 8 was identical to the procedure for cycle 6. The teams managed the emergency posed by C3Fire scenario H.

Thank you and goodbye

When Team 2 had completed its review of its play in scenario H, all eight participants were brought together for the last time. Their unstructured conversation was often filled with recommendations and ideas for future experiments. Participants were paid individually, thanked for their participation, and encouraged to tell their friends about the opportunity to take part in the experiment.

Plans for the Second Half of 2005 and 2006

In the second half of 2005, we plan to (A) conduct additional experiments with foreign nationals who have recently moved to Sweden, (B) analyze the data from these experiments, (C) submit at least one article for publication in a peer-reviewed journal, and (D) write a formal proposal for funding to ARO to continue our lines of research beyond the end-date of the MURI OPUS grant in 2006.

Data analysis

After the briefest of summer vacations, we will all start work on the quantitative and qualitative analyses of the behavioral and verbal data. The experimental paradigm produces a mountain of data that needs to be approached in a systematic and disciplined manner. I anticipate that the interim report for December 2005 will detail this approach in much the same way as this report details the experimental paradigm.

Experiments with foreign nationals

It has not been easy to solicit the participation of groups of eight foreign-born adult men with Swedish residence permits. We have made progress with the local Bosnian and Italian communities but both were reluctant to commit to spending two blocks of four hours of their free time with us during the spring. We hope that this attitude will change with the change in seasons and that we will find it easier to recruit participants in the fall.

While Bosnia and Italy are both Mediterranean countries, there is good reason to expect that their cultural norms for decision making will vary. Figure 1 is a plot from Hofstede (2001) showing two cultural dimensions along which these neighbors vary. The horizontal axis is Hofstede's dimension of 'Power Distance' defined as comfort with differential authority. The vertical axis is Hofstede's dimension of 'Individualism / Collectivism,' a measure of the span of the locus of concern. The axes mark the median values for Hofstede's sample of 53 cultures.

The three cultures in the lower left corner (S – Sweden, U – United States, and It – Italy) are individualistic and relatively intolerant of large difference in authority. In contrast, Yugoslavians (Y, Bosnians?) and Filipinos (P) plot in the upper right-hand corner. These cultures are relatively comfortable with highly discrepant levels of power (i.e., they tolerated dictatorships by Tito and Marcos) and are more committed to the good of the group than the prosperity of the individual.

We have made preliminary contact with the Iranian student association at the University and hope to recruit several groups of Farsi-speakers. It is interesting to note that Iranians (Ir) plot near the median in Figure 1. This strong central tendency for Iranians is seen in all of Hofstede's dimensions.

Our contacts at Migrationsverket and at Linköping University have promised to help us get in contact with an interesting mix of foreign nationals. We are told there are communities of Burundians (black, French speakers) and Tagalog (pre-Spanish conquest, native Filipinos) near Linköping. We do not know where these cultures map on Hostede's dimensions as they come from cultures and parts of the world where ethnography is rare.

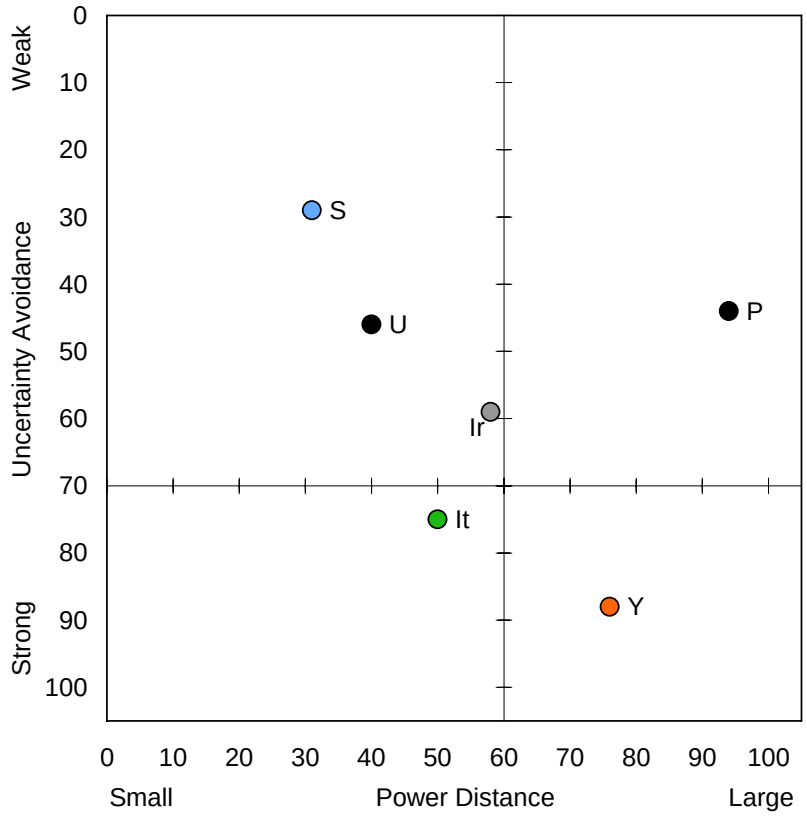


Figure 1. Graph from Hofstede (2001) showing how six cultures vary along two dimensions of cultural difference, Power Distance and Collectivism. Culture codes: S – Sweden, U – United States, It – Italy, Ir – Iran, Y – Yugoslavia, P – Philippines. Hofstede's data were collected prior to the dissolution of Yugoslavia.

Article to peer reviewed journal

I will be presenting a poster on our experimental paradigm at the 7th European conference on Cross Cultural Psychology in San Sebastian Spain, July 11 to 15, 2005. At the meeting, I hope to meet new colleagues in the community of scientists studying allied issues. One key piece of information I hope to obtain is a short list of respected journals where our research would be treated to both sympathetic and critical peer-review.

Proposal to ARO for continued funding beyond the lapse of the MURI-OPUS grant

We want our work to evolve in a manner that is concordant with ARO's needs and goals. The only way for us to be confident that this is the case is to work with ARO in the generation of a proposal for continued funding. We look forward to hearing from you and encourage you to visit us in Linköping when you have the time.

References

Artman, H. (1999). *Fördelade kunskapsprocesser i ledningscentraler vid nödsituationer.* Unpublished PhD Thesis, Linköpings Universitet, Linköping.

Ball, S. A., Rounsaville, B. J., Tennen, H., Kranzler, H. R. (2001). Reliability of Personality Disorder Symptoms and Personality Traits in Substance-Dependent Inpatients. *Journal of Abnormal Psychology, 110 (2),* 341-352.

Barker, J., Tjosvold, D., & Andrews, I. R. (1988). Conflict approaches of effective and ineffective managers: A field study in a matrix organization. *Journal of Management Studies, 25,* 167-178.

Bar-Tal, Y. (1994). The effect on mundane decision-making of the need and ability to achieve cognitive structure. *European Journal of Personality, 8,* 45-53.

Brehmer, B. (2005) Micro-worlds and the circular relation between people and their environment, *Theoretical Issues in Ergonomics Science*, 6, 1, 73-93.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior, 9*(2-3), 171.

Costa, P.,T., Jr and McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor (NEO-FFI) Inventory professional manual.* Odessa, FL: PAR Inc.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper Perennial.

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. New York: Basic Books.

Ericsson, K. A., and Simon, H. A. (1984/1995). *Protocol Analysis: Verbal Reports as Data*. Cambridge, Mass.: MIT Press.

Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. In G. Strube & K. F. Wender (Eds.), *The Cognitive Psychology of Knowledge*. Amsterdam: Elsevier.

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning,* 7(1), 69-89.

Granlund, R. (2002). *Monitoring Distributed Teamwork Training.* Unpublished PhD Thesis, Linköpings Universitet, Linköping, SE.

Granlund, R. (2003). Monitoring experiences from command and control research with the C3Fire microworld. *Cognition, Technology & Work, 5*(3), 183-190.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly, 2*, 205-227.

Greco, V. & Roger, D. (2001). Coping with uncertainty: The construction and validation of a new measure. *Personality and Individual Differences*, 31, 519-534.

Hofstede, G. (2001). *Culture's consequences: Second Edition*. Thousand Oaks, CA: Sage.

Jackson, S. A. & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport and Exercise Psychology, 18,* 17-35.

Johansson, B. (2003). *Feedforward Control in Dynamic Situations.* Unpublished Licentiate Thesis, Linköpings Universitet, Linköping, Sweden.

Johansson, B., Granlund, R. & Waern, Y (2004). Research on Decision Making and New Technology - Methodological Issues. In (Eds.) B. Brehmer, R. Lipshitz, & H. Montgomery, *How Professionals Make Expert Decisions*. Mahaw, New Jersey: Lawrence Erlbaum Associates.

Johansson, B., Persson, M., Granlund, R., & Mattsson, P. (2003). C3Fire in command and control research. *Cognition, Technology & Work, 5*(3), 191-196.

Kluckhohn, F. & Strodtbeck, F. (1961). *Variations in value orientations*. Evanston, Illinois: Row, Peterson and Company.

Kluckhohn Center. (1995). *User's manual for the Value Orientation Method*. Kluckhohn Center for the Study of Values: Bellingham, WA.

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224-253.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.

Novy D. M., Nelson D. V., & Goodwin J. (1993). Psychometric comparability of the Stait-Trait Anxiety Inventory for different ethnic subpopulations. *Psychological Assessment, 5,* 343-349.

Rigas, G., Carling, E., & Brehmer, B. (2002). Reliability and validity of performance measures in microworlds. *Intelligence, 30*, 463-480.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y).* Palo Alto, CA: Consulting Psychologist Press.

Svenmarck, P. (1998). *Local Co-ordination in Dynamic Environments: Theories and Co-ordination Support.* Unpublished Licentiate Thesis, Linköping University: Linköping, Sweden.

Tjosvold, D. (1985). Implications of controversy research for management. *Journal of Management, 11,* 21-37.

Triandis, H. C. (1994). *Culture and social behavior*. New York: McGraw-Hill.

Adaptive Change in the Type of Automation Support Reduces the Cost of Imperfect Decision Aids in a Simulated Battlefield Engagement Task

Kathleen McGarry

Raja Parasuraman



Cognitive Science Laboratory, Catholic University of America

GEORGE MASON UNIVERSITY

Introduction

- Current military operations involve a degree of complexity that mandates use of automated decision-making tools to assist human operator
- Automation can lead to both benefits and costs in human and system performance

Introduction, cont.

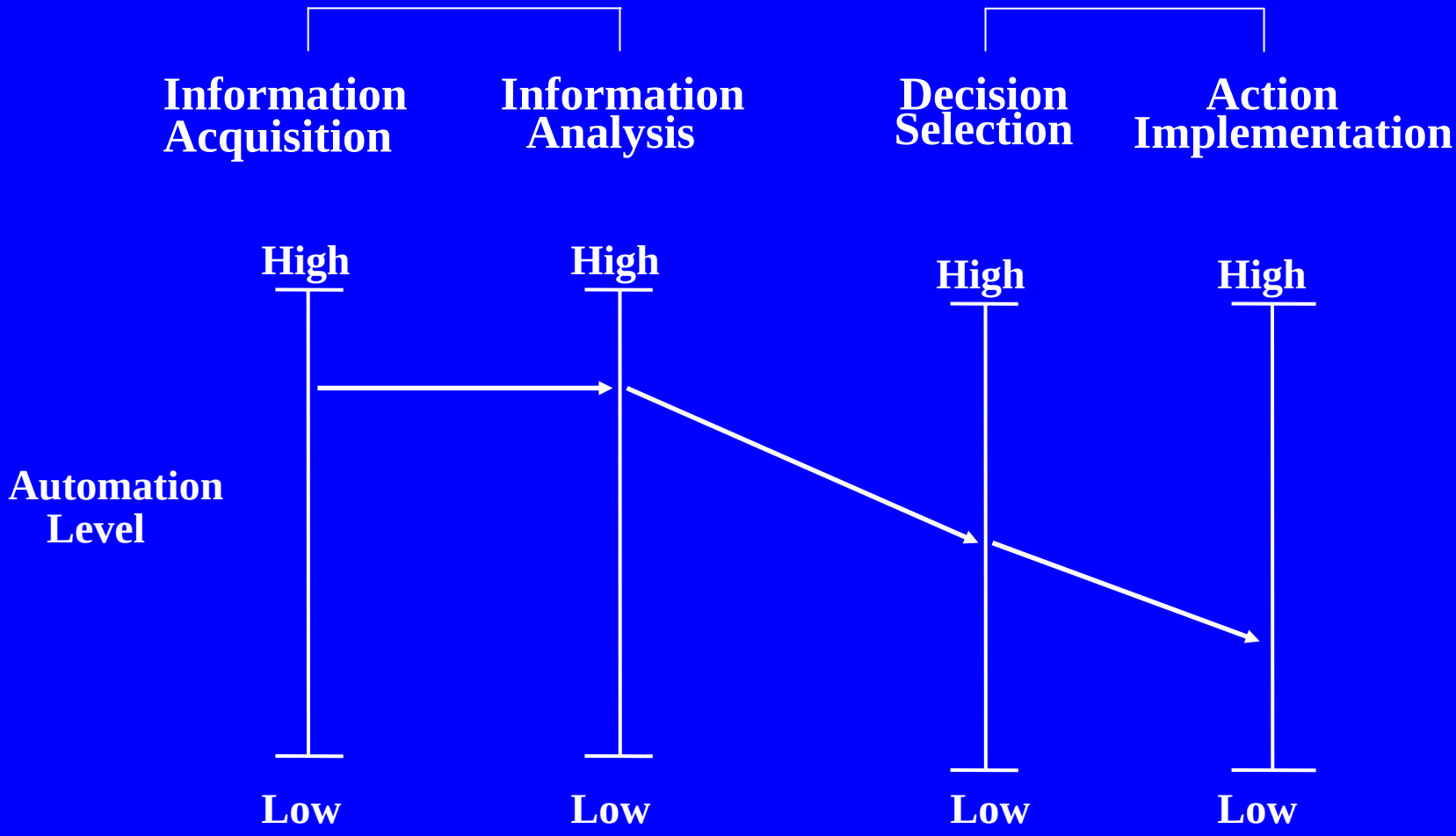
- Automation generally enhances performance, reduces operator workload, improves system efficiency
- Incurred costs
  - Loss of coordination between human and automated agents
  - Over-reliance
  - Mistrust
  - Loss of operator skill
  - Unbalanced workload



Introduction, cont.

- Automation can support the battlefield commander
  - Collect and present multiple information sources in organized manner
  - Assist human operator in high-stress, high-risk environments
  - Complete information analysis and present best course of action
  - Action itself can be automated

Information Automation Decision Automation



(Adapted from Parasuraman, Sheridan, & Wickens, 2000)

Introduction, cont.

- Crocoll & Coury (1990) - air defense task

  -information automation (status) - identifies an aircraft

  -decision automation (command) - “fire” or “no fire”

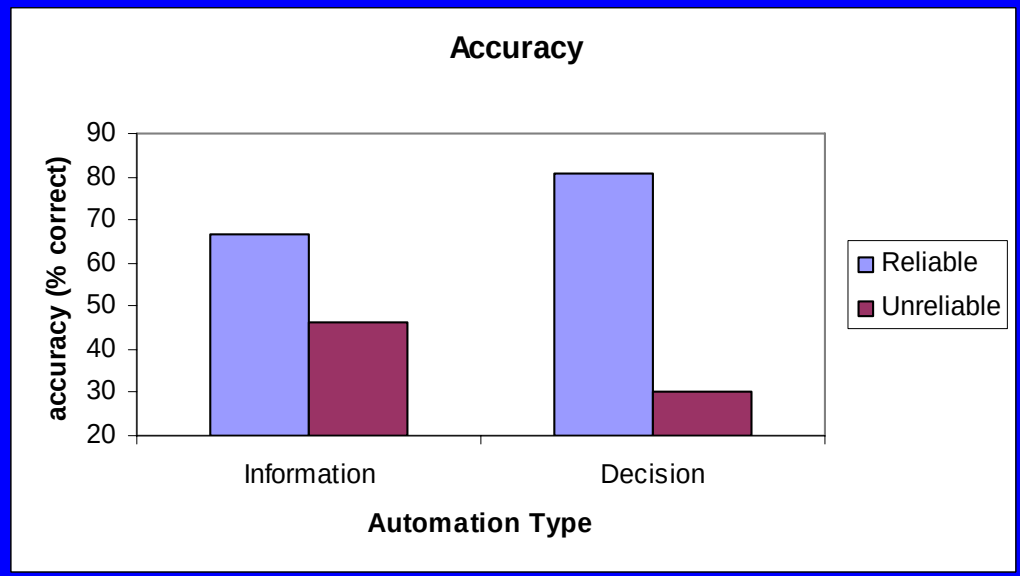
- Results - unreliable automation

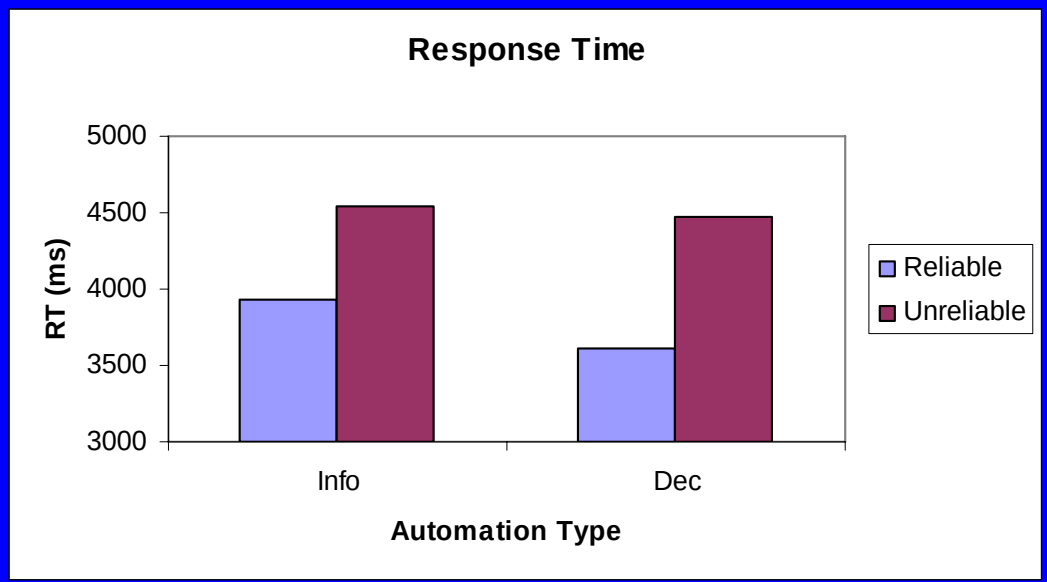
  -greater detrimental effect on performance when it provided decision automation than information automation

Introduction, cont.

- McGarry, Rovira & Parasuraman (2003): battlefield task
  - Information automation: complete listing
  - Decision automation: recommendation
- Results:
  - Reliable decision automation significantly reduced response time as compared to information automation
  - Benefit counteracted by slower and less accurate performance when decision automation was unreliable

McGarry, et al (2003) Results





Introduction, cont.

- **Parasuraman, Mouloua & Molloy (1996)**

  - MAT task

  -Adaptive automation
  - Model-based
  - Performance-based

  -Static automation

- Results

  -Performance improved after return to manual control

Research Questions

- Can the costs associated with imperfect automation be mitigated?
- Will an adaptive change in automation allow an individual to build up a mental model of the automation, leading to a reduction in cost when automation is unreliable?
- Does type of adaptive automation have an effect on an individuals' performance?

Design

Condition

Static Automation |---------| |---------| |---------|

Dec Dec Dec

Adaptive Automation

Information |---------| |---------|

Dec |---------| Dec

Info

Manual |---------| |---------|

Dec |---------| Dec

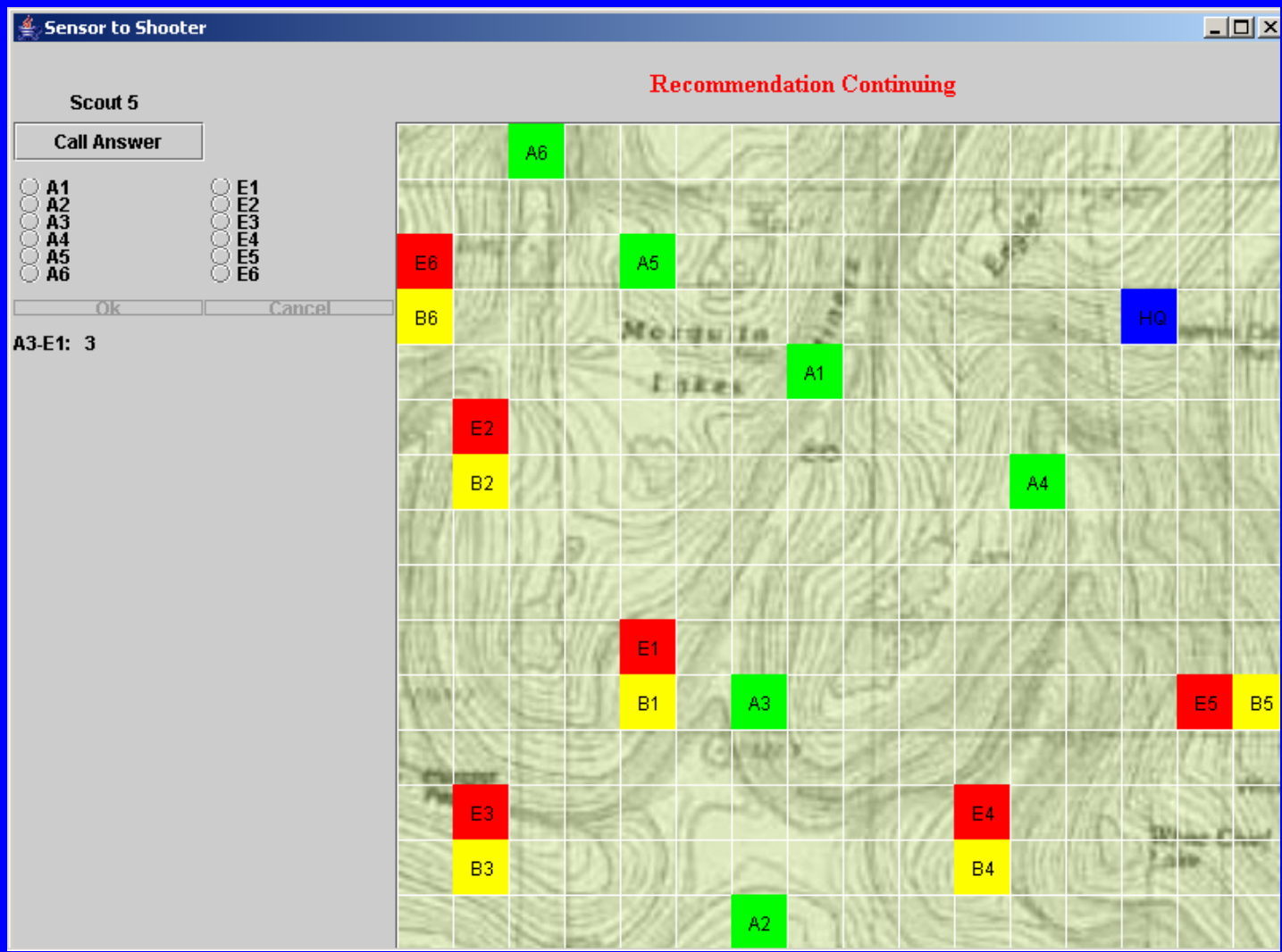
Manual

10 min 20 min 30 min

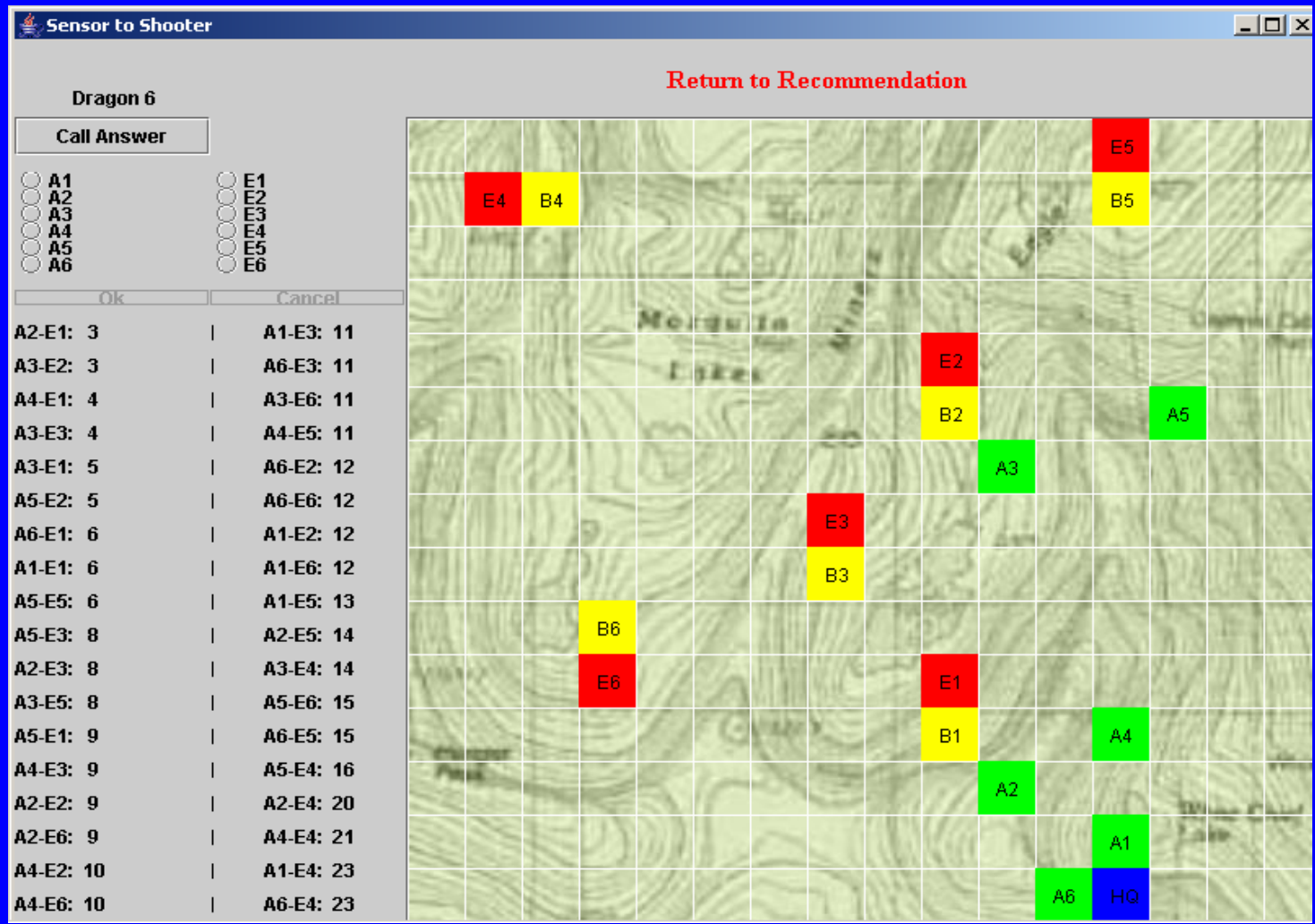
Methods

- Participants
  - 12 paid volunteers
    - 8 F, 4 M
    - Ages 18-22; M = 19.5
- Design
  - Within Subjects
  - 3 conditions
    - Static Automation
    - Information Adaptive Automation
    - Manual Adaptive Automation
- Dependent Variables
  - Decision accuracy
  - Decision response time

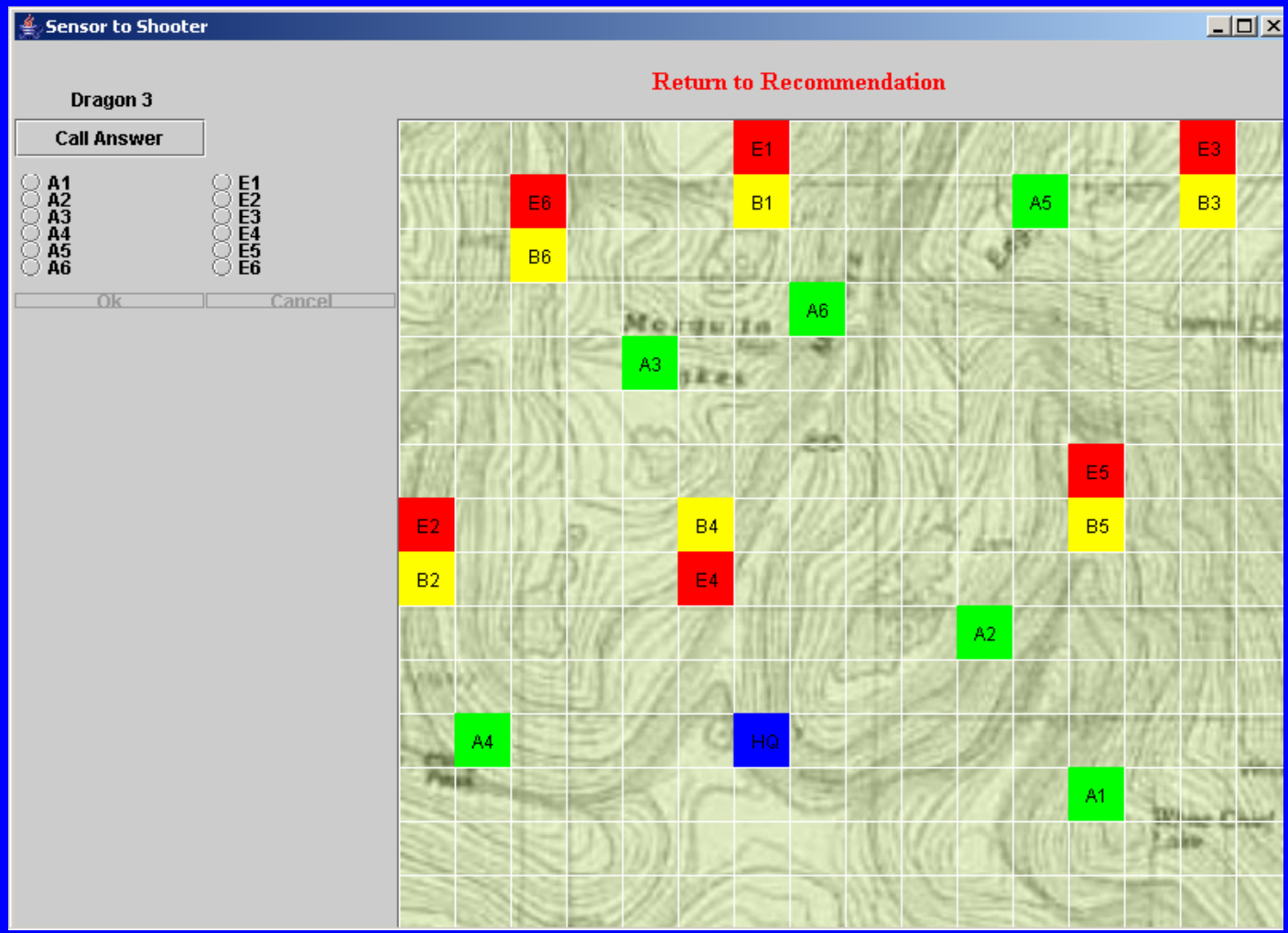
Static Automation

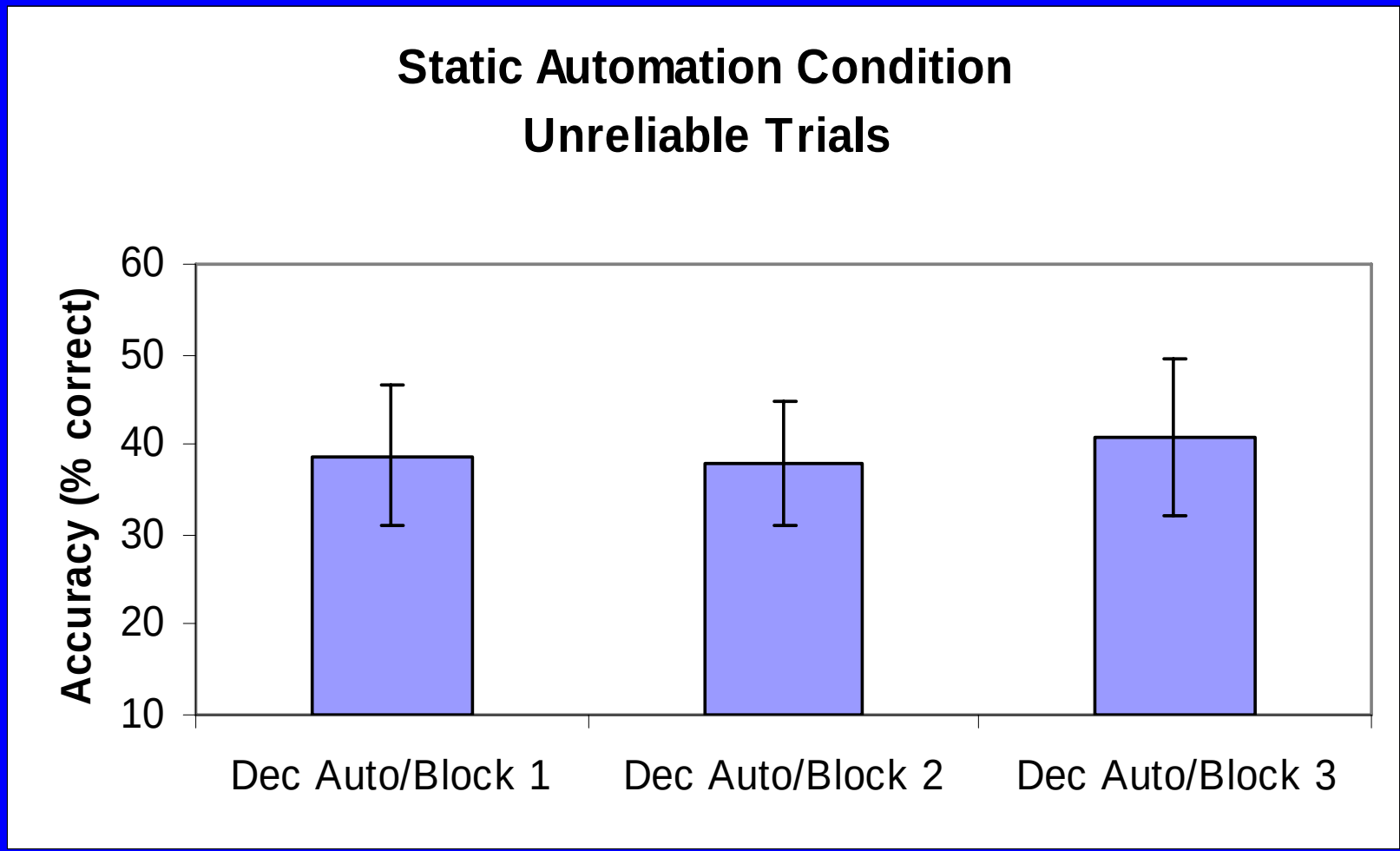


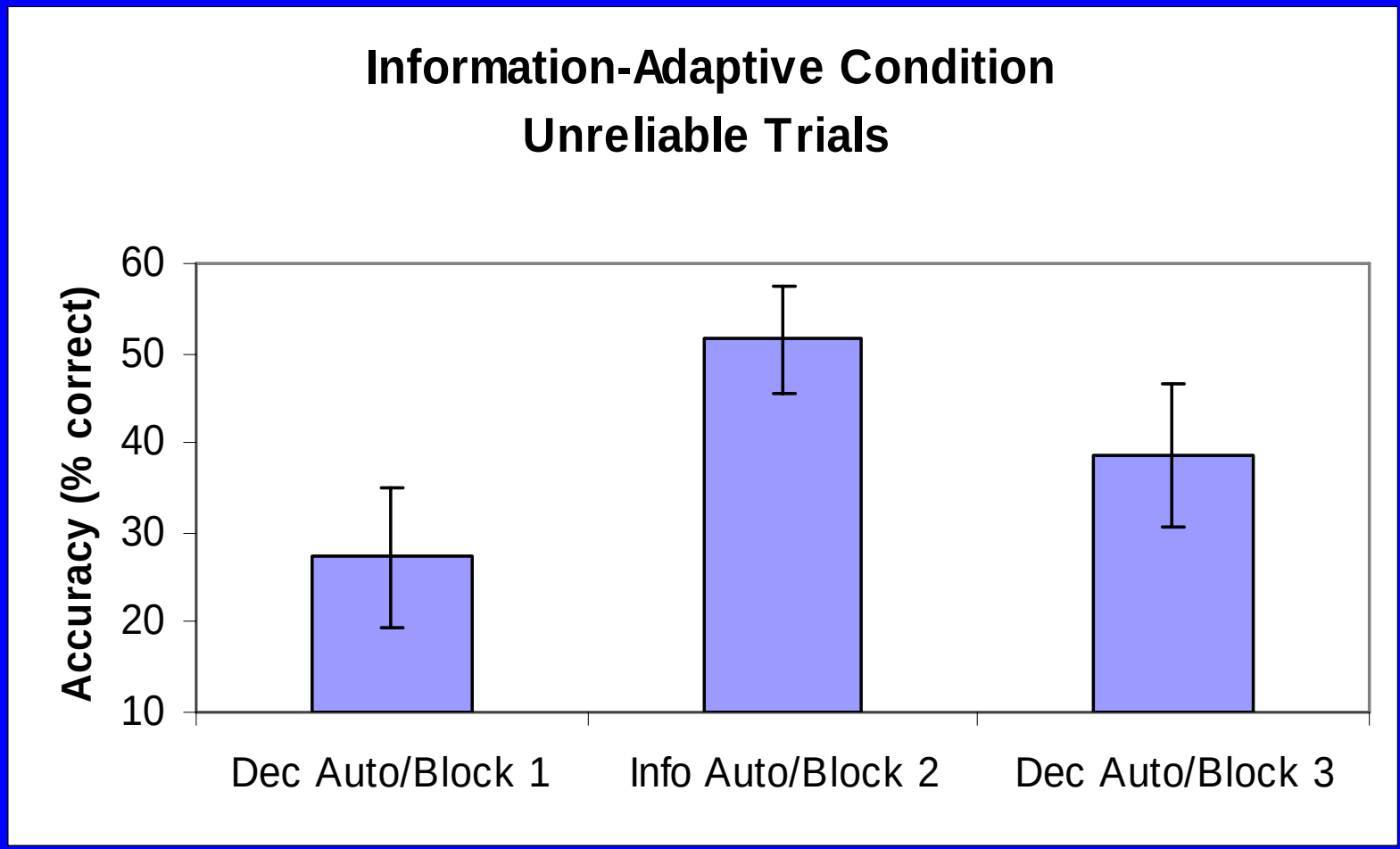
Information Adaptive Automation



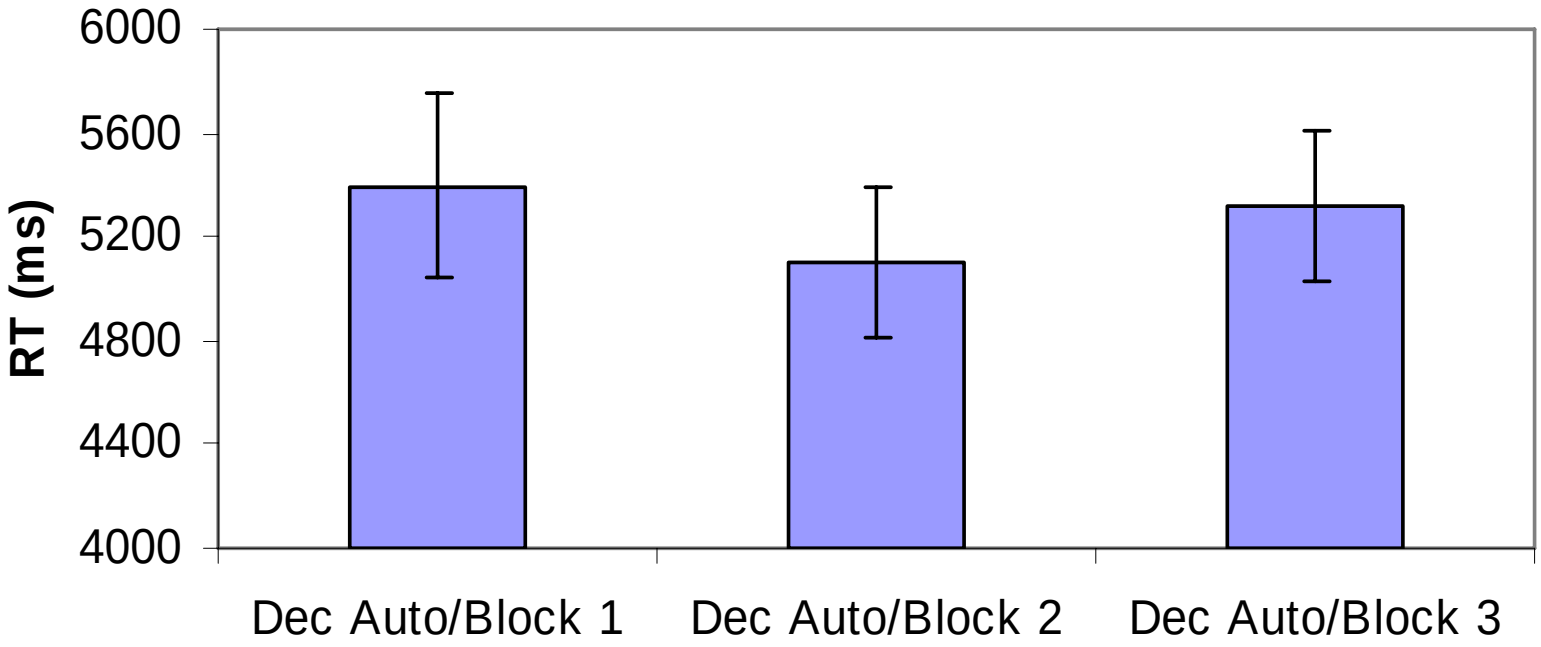
Manual Adaptive Automation

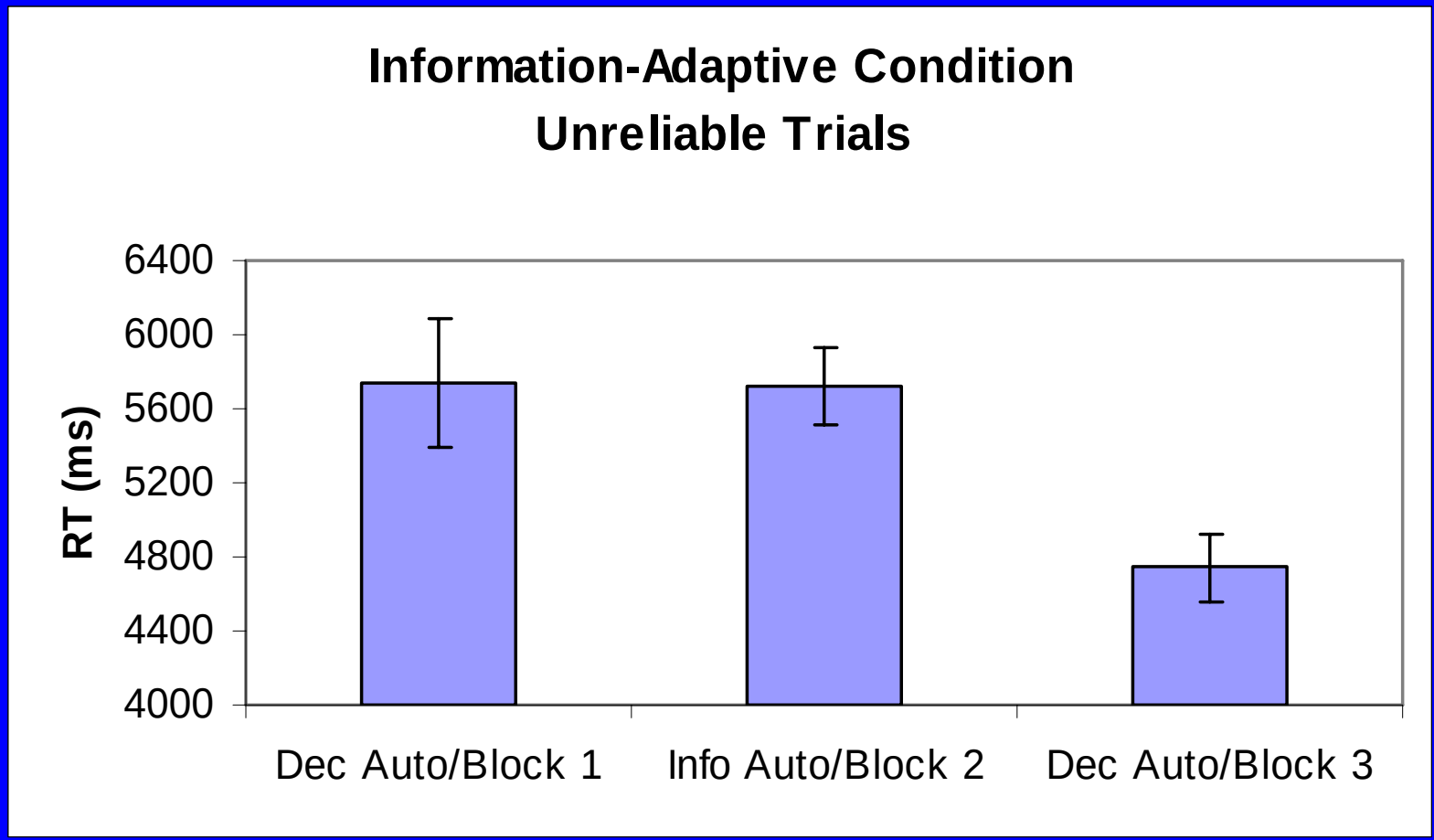


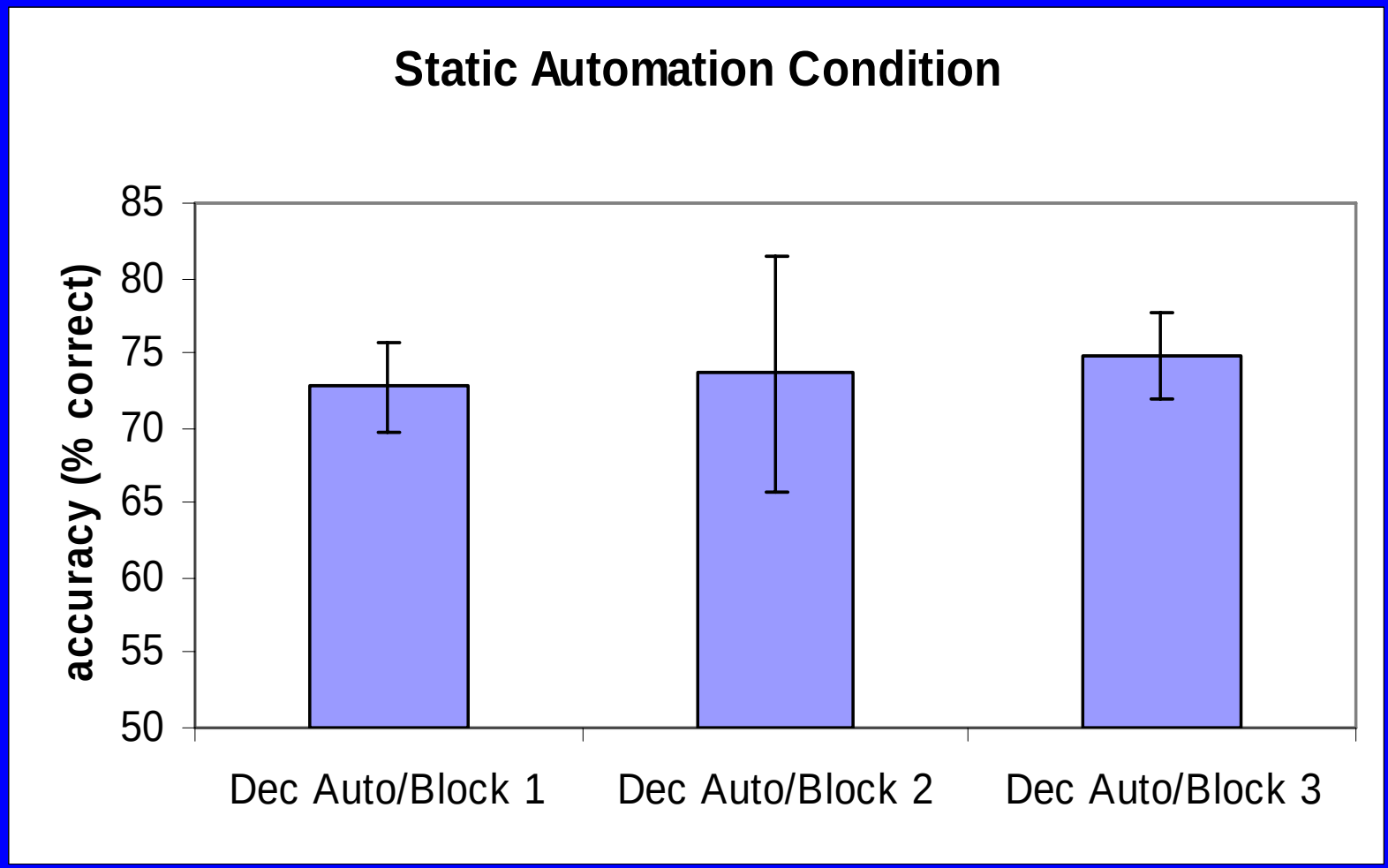

Static Automation Condition
Unreliable Trials
Accuracy (% correct)
60
50
40
30
20
10
Dec Auto/Block 1
Dec Auto/Block 2
Dec Auto/Block 3

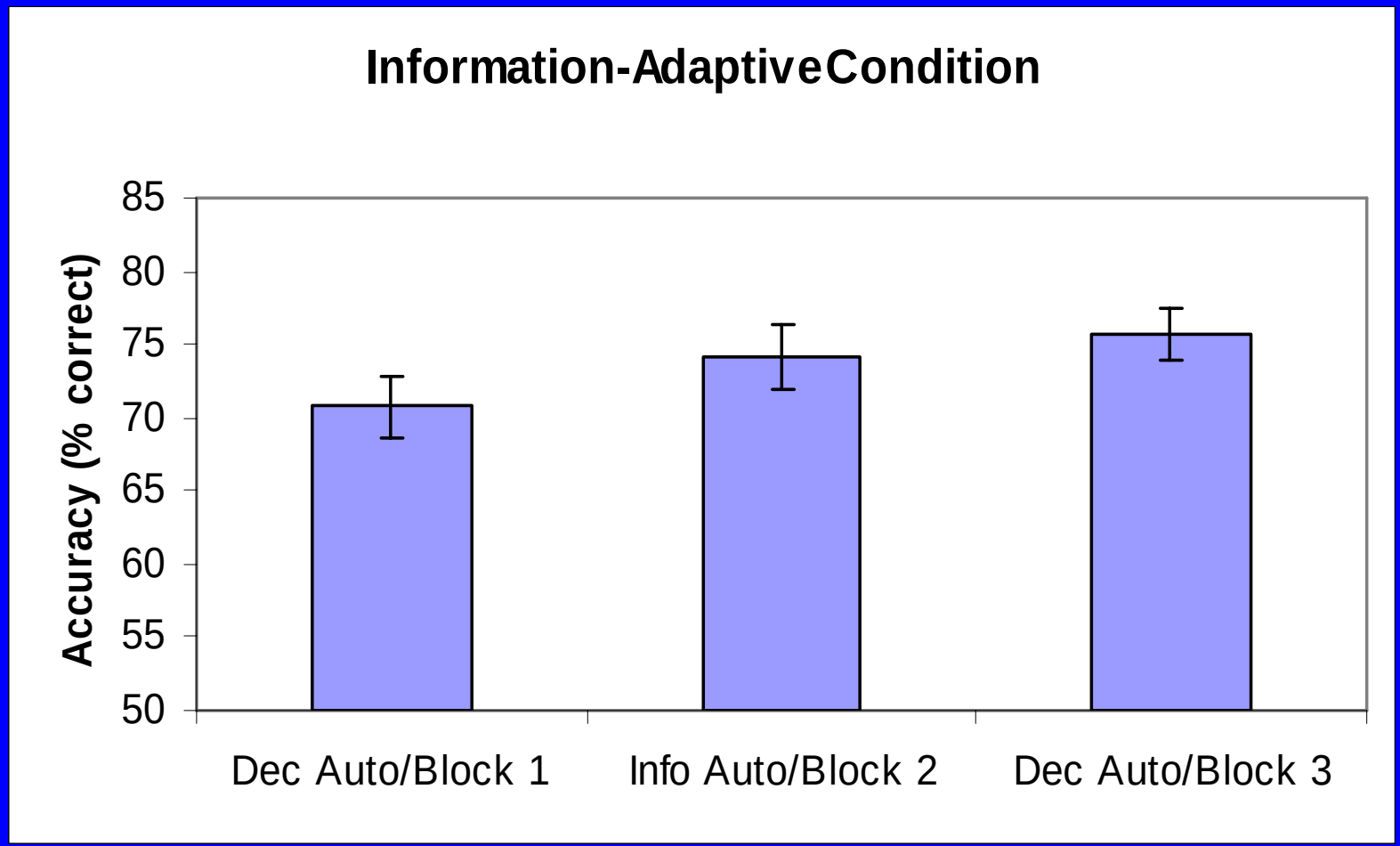

Information-Adaptive Condition
Unreliable Trials
Accuracy (% correct)
60
50
40
30
20
10
Dec Auto/Block 1
Info Auto/Block 2
Dec Auto/Block 3

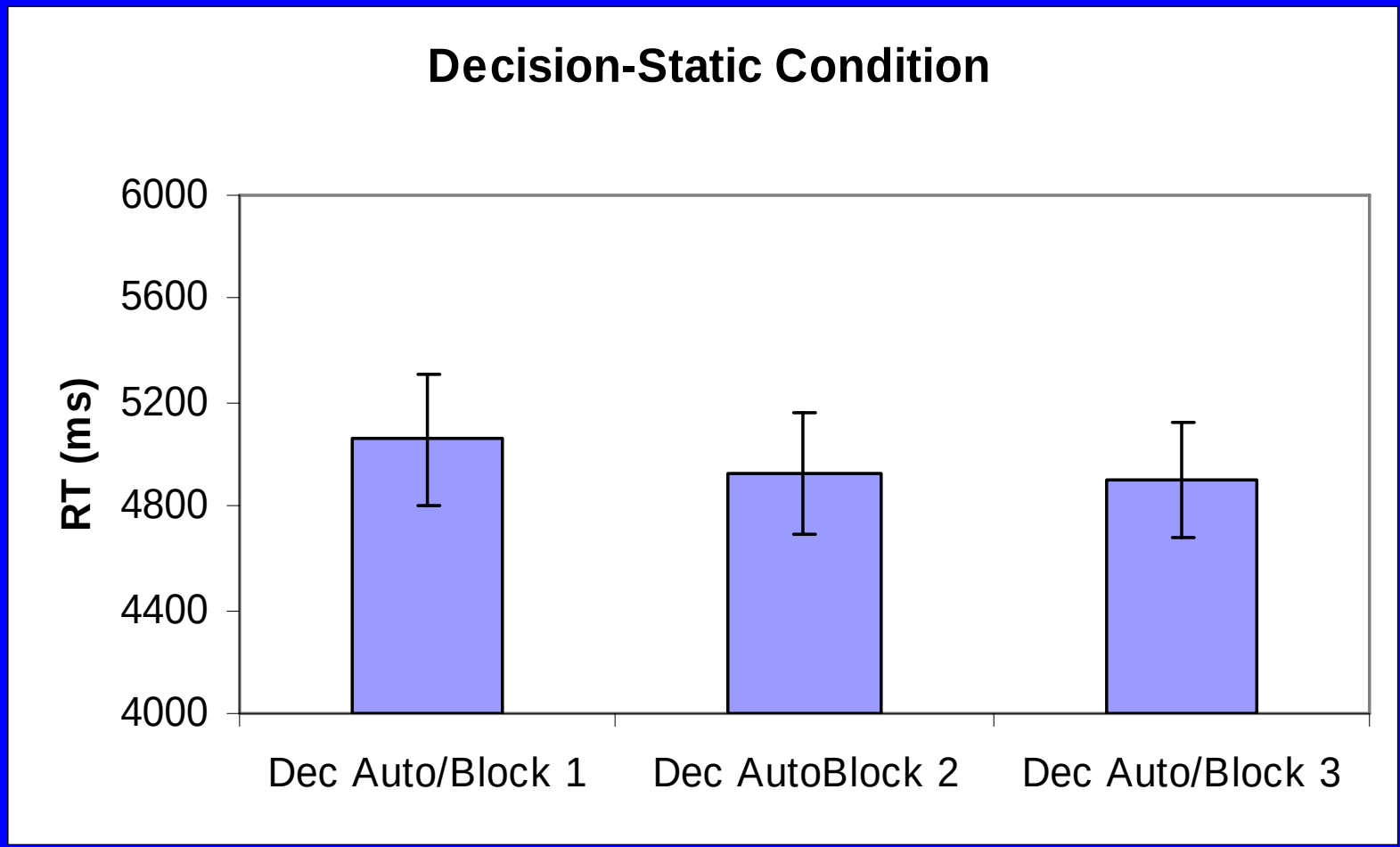
Static Automation Condition
Unreliable Trials

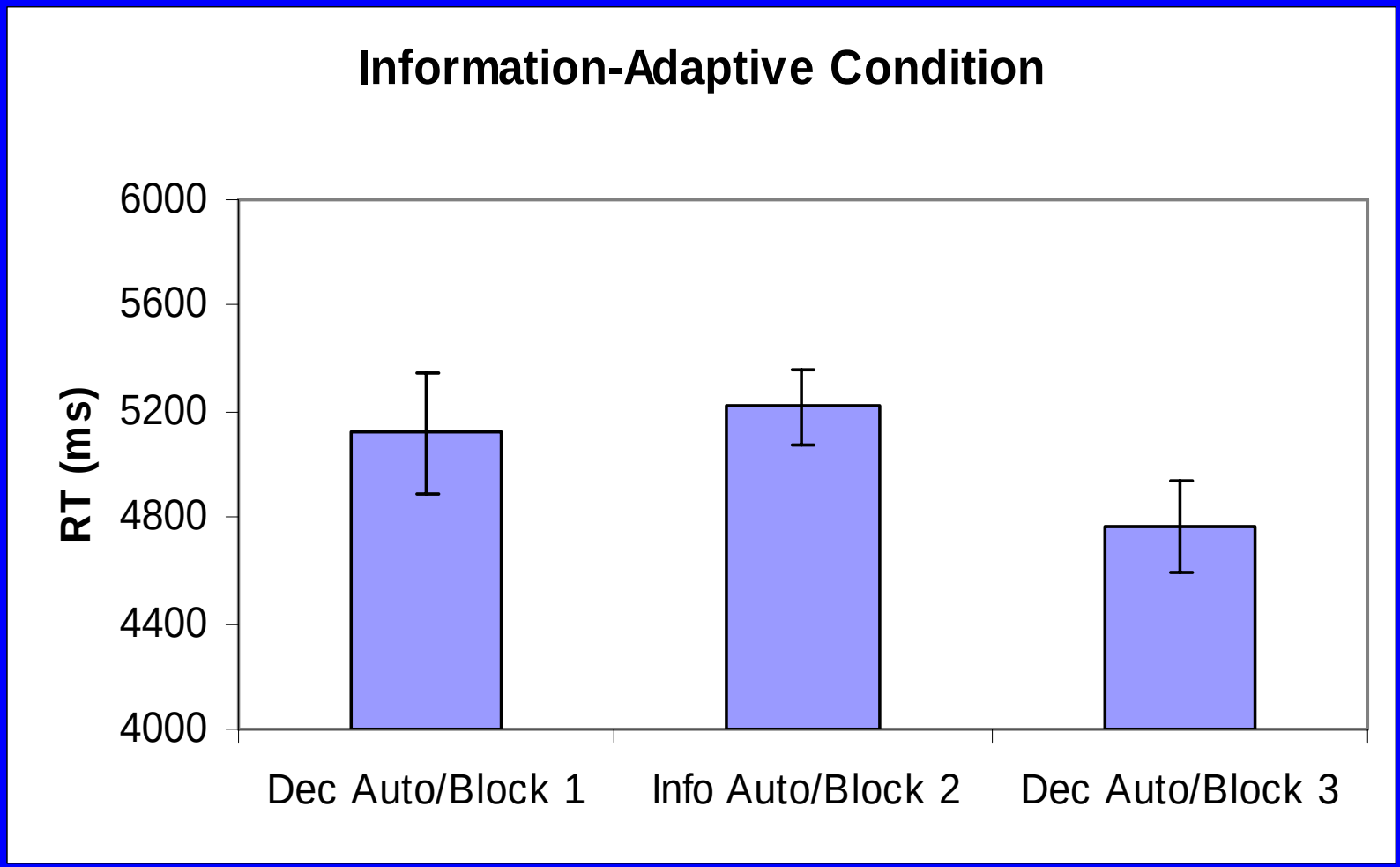


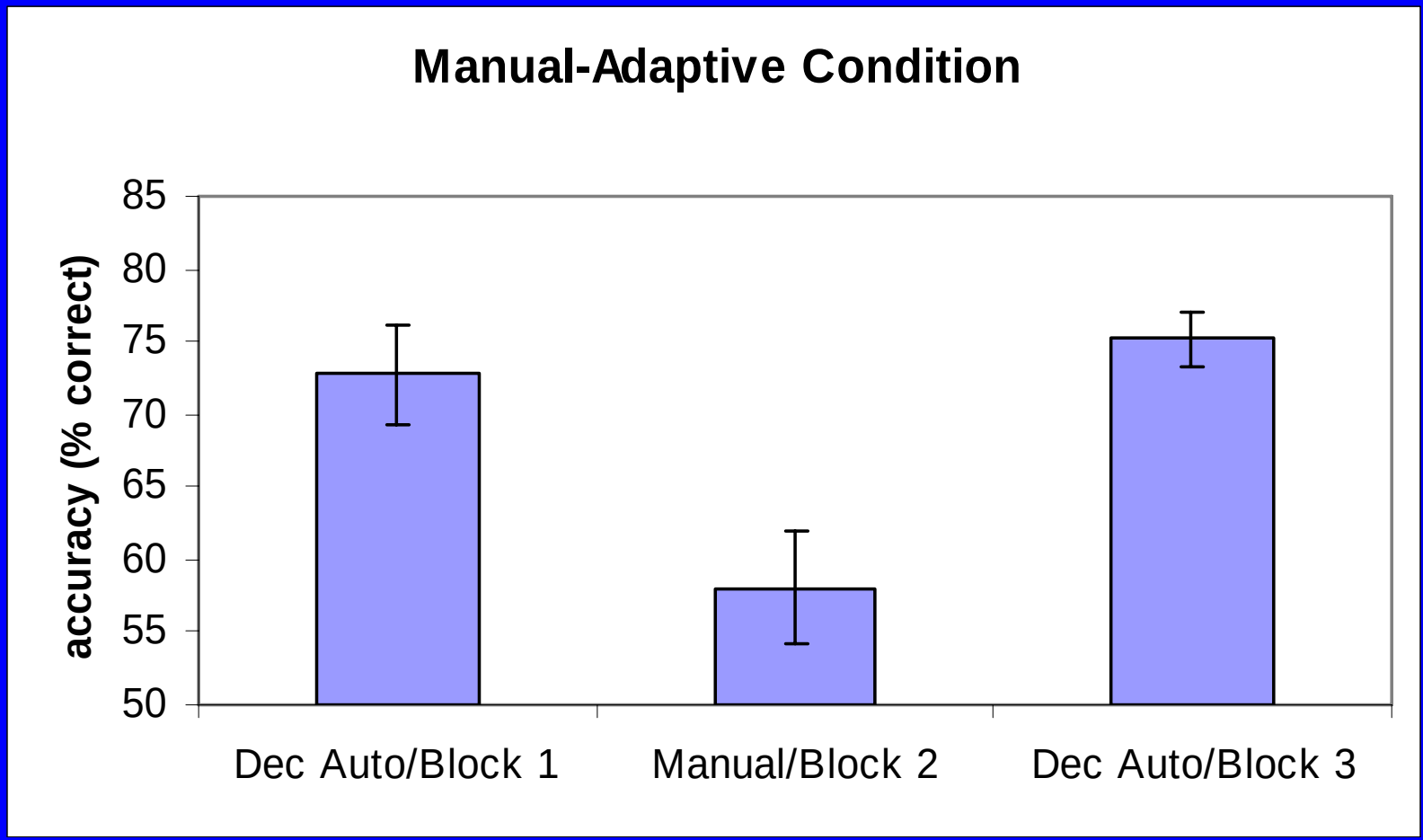

Information-Adaptive Condition
Unreliable Trials
RT (ms)
6400
6000
5600
5200
4800
4400
4000
Dec Auto/Block 1
Info Auto/Block 2
Dec Auto/Block 3

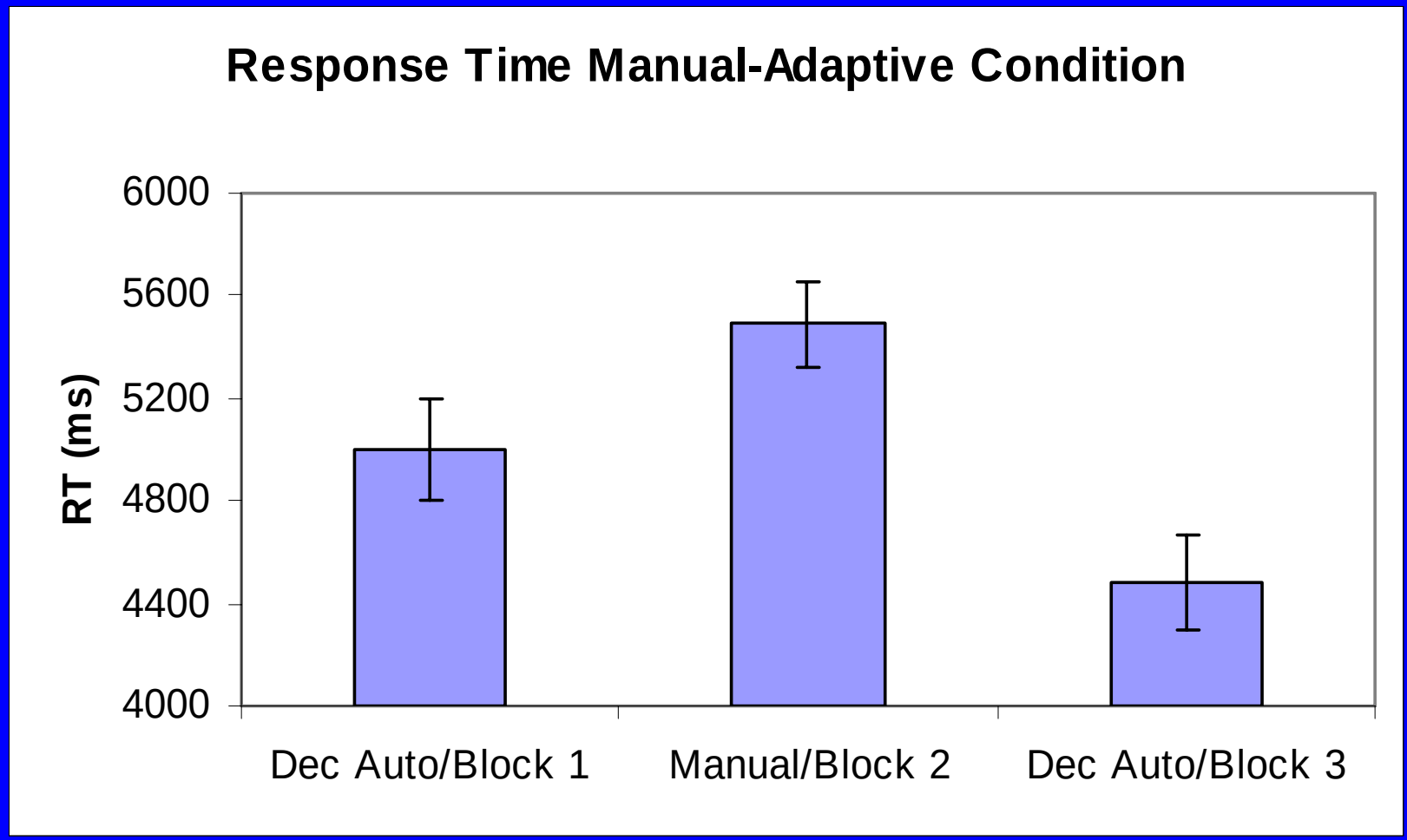

Static Automation Condition
accuracy (% correct)
85
80
75
70
65
60
55
50
Dec Auto/Block 1
Dec Auto/Block 2
Dec Auto/Block 3


Information-AdaptiveCondition
Accuracy (% correct)
85
80
75
70
65
60
55
50
Dec Auto/Block 1
Info Auto/Block 2
Dec Auto/Block 3


Decision-Static Condition
RT (ms)
6000
5600
5200
4800
4400
4000
Dec Auto/Block 1
Dec AutoBlock 2
Dec Auto/Block 3


Information-Adaptive Condition
RT (ms)
6000
5600
5200
4800
4400
4000
Dec Auto/Block 1
Info Auto/Block 2
Dec Auto/Block 3


Manual-Adaptive Condition
accuracy (% correct)
85
80
75
70
65
60
55
50
Dec Auto/Block 1
Manual/Block 2
Dec Auto/Block 3


Response Time Manual-Adaptive Condition
RT (ms)
6000
5600
5200
4800
4400
4000
Dec Auto/Block 1
Manual/Block 2
Dec Auto/Block 3

Summary of Results

- Overall performance improved after an adaptive change in automation to information automation
  - RT performance also improved after return to manual control
- Cost was lessened with change in automation type in middle of task
  - Both accuracy and RT improved with change to information automation
  - RT improved with return to manual control

Conclusions

- Adaptive change in automation lessens cost of unreliable automation
- Adaptive change in type of automation improves performance
  - When change is from recommendation to listing, both accuracy and RT improves
  - When change is from decision recommendation to return-to-manual, RT improves

Acknowledgments

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. The views expressed in this work are those of the author and do not necessarily reflect official Army policy. We would also like to thank Tashfeen Bhimdi for his technical support.

Interim Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield
Army MURI, Subcontract to U. Central Florida

Raja Parasuraman
George Mason University

PERFORMANCE PERIOD: 1/1/05-4/30/05

The following study was completed during this performance period.

Effects of Adaptive Decision Support in a Simulated Battlefield Engagement Task

This study follows on from previous studies under the MURI on the use of adaptive automation to mitigate stress, workload, and fatigue effects in a simulated battlefield engagement task. Twelve participants were provided with two types of automation (decision and information) and also performed the task manually. In three conditions, the type of automation was alternated during performance of the task over three blocks of trials. In all three conditions, decision automation was provided in the first and the third blocks of the task, with the middle block requiring the use of decision automation, information automation, or manual performance. The accuracy of engagement decisions improved in the third block with decision automation when it was preceded by a temporary adaptive change to information automation. No such improvement occurred when decision automation was used throughout the task or when the adaptive change involved a temporary return to manual performance. This suggests that providing the user with short periods of information automation can help mitigate some of the costs of imperfect decision automation by keeping the operator in the decision-making loop. A computational analysis was conducted to determine the optimal cycling periods of information and decision automation required to achieve a given level of system performance.

The study was presented by Katie McGarry at the Annual Mid-Year Meeting of Division 21 of the American Psychological Association, Fairfax, VA, March 3-4, 2005.

STANDARD

NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION

AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY
RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS
(MURI / OPUS) LABORATORY

AGREEMENT TITLE: Operator Under Stress Research

AGREEMENT NUMBER: NCRADA-**NAVAIR ORLANDO TSD-02-037**

AGREEMENT ADMINISTRATORS:

NAVAIR ORLANDO TSD

Technology Transfer Office:	**Ms. Patricia R. Nimmo, AIR 49T, (407) 380-8225**
Legal Counsel:	**Robert W. Adams, AIR 772P, (407) 380-8211**
Principal Investigator:	**Ron Wolff, AIR 4962, (407) 380-4583**

UCF MURI / OPUS

Preferred Contact:	**Mrs. Kim Smith, UCF Office of Research, (407) 823-3062**
Principal Investigator:	**Dr. Peter Hancock, Director, UCF MURI/OPUS Laboratory, (407) 823-2310**

This Page Intentionally Left Blank

TABLE OF CONTENTS

PREAMBLE

Article 1.

DEFINITIONS 2

1.1 "Agreement" 2
1.2 "Classified Information" 2
1.3 "Collaborator" 2
1.4 "Controlled Unclassified Information (CUI)" 2
1.5 "Cooperative Work" 2
1.6 "Data" 2
1.7 "Effective Date" 2
1.8 "Exclusive License" 2
1.9 "Government" 2
1.10 "Government Purpose Rights" 2
1.11 "Information" 2
1.12 "Intellectual Property" 2
1.13 "Invention" 2
1.14 "Invention Disclosure" 3
1.15 "Made" 3
1.16 "Militarily Critical Technologies (MCT)" 3
1.17 "Non-Subject Data" 3
1.18 "Non-Subject Invention" 3
1.19 "Patent Application" 3
1.20 "Principal Investigator (PI)" 3
1.21 "Proprietary Information" 3
1.22 "Restricted Access Information" 3
1.23 "Subject Data" 3
1.24 "Subject Invention" 3
1.25 "Tangible Property" 4
1.26 "Unlimited Rights" 4

Article 2. OBJECTIVES 4

Article 3. RESPONSIBILITIES 4

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities 4
3.2 **UCF MURI / OPUS Laboratory** Personnel and facilities 5
3.3 Security Regulations and Directives 5

Article 4. REPRESENTATIONS AND WARRANTIES 5

4.1 **NAVAIR ORLANDO TSD** Representations and Warranties 5
4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Representations and Warranties 5
4.3 Joint Representations 6

Article 5. FUNDING 7

5.1 Payment Schedule..... 7
5.2 Insufficient and Excess Funds..... 7
5.3 No New Commitments..... 8
5.4 Accounting Records..... 8

Article 6. REPORTS AND PUBLICATIONS..... 8

6.1 Final Reports..... 8
6.2 Agreement to Confer Prior to Publication or Public Disclosure..... 9
6.3 Classified Information..... 9

Article 7. INTELLECTUAL PROPERTY..... 9

7.1 Data..... 9
7.1.1 General Provisions Applying to All Data..... 9
7.1.1.1 Ownership..... 9
7.1.1.2 No Implied License..... 9
7.1.1.3 Marking of Data..... 10
7.1.1.3.1 Data Provided With Less Than Unlimited Rights.. 10
7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information..... 10
7.1.1.3.3 Data That are Subject to 35 USC 205..... 11
7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted..... 11
7.1.1.4 Protection of Data..... 11
7.1.1.5 Release of Data Under the Freedom of Information Act..... 11
7.1.2 Subject Data..... 12
7.1.2.1 Delivery of Requested Subject Data..... 12
7.1.2.2 Rights in Subject Data..... 12
7.1.3 Rights in Non-Subject Data..... 12
7.2 Copyrights..... 13
7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory**..... 13
7.2.2 Copyright License to the Government..... 13
7.2.3 Copyright Statement..... 13
7.3 Trademarks and Service Marks..... 13
7.3.1 Ownership of Trademarks and Service Marks..... 13
7.3.2 Obligation of Employees to Report Trademarks and Service Marks.... 13
7.3.3 Obligation of Collaborators to Notify Each Other..... 13
7.3.4 Responsibility for Filing an Application for Trademark or Service Mark..... 14
7.3.5 License to Use Trademark or Service Mark..... 14
7.4 Subject Inventions..... 14
7.4.1 Obligation to Report Subject Inventions..... 14
7.4.1.1 Collaborators' Instruction to Employees..... 14
7.4.1.2 Timely Invention Disclosure by Inventors..... 14
7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator..... 14
7.4.2 Determination of Subject Inventions..... 14
7.4.3 Title to and Ownership of Subject Inventions..... 14

7.4.4 Filing of Patent Applications..... 15
7.4.4.1 Filing of Patent Applications on Solely Made Inventions..... 15
7.4.4.2 Filing of Patent Applications on Jointly Made Inventions..... 15
7.4.4.3 Preserving Intellectual Property Rights..... 15
7.4.4.4 Filing Deadlines..... 15
7.4.4.5 Copies and Inspection..... 15
7.4.4.5.1 Copies of Prosecution Papers..... 15
7.4.4.5.2 Access to Patent Application File and Right to Make Copies..... 16
7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application..... 16
7.4.5 Nonexclusive License to Subject Inventions..... 16
7.4.5.1 Nonexclusive License Grant..... 16
7.4.5.2 Confirmatory Nonexclusive License Agreement..... 16
7.4.6 Option for Exclusive License to Subject Inventions..... 16
7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement. 17
7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions..... 17
7.4.8.1 Exclusive Licenses and Exclusive License Option..... 17
7.4.8.2 Nonexclusive Licenses..... 17
7.5 Non-Subject Inventions..... 18
7.5.1 Ownership of Non-Subject Inventions..... 18
7.5.2 Rights Under Other Agreements..... 18
7.5.3 No License to Non-Subject Inventions..... 18
7.6 Research License..... 18

Article 8. TANGIBLE PROPERTY..... 18

8.1 Title to Preexisting Tangible Property..... 18
8.2 Tangible Property Purchased by Collaborators to Perform Cooperative Work... 18
8.3 Title to Developed Tangible Property..... 19
8.4 Tangible Property Operational and Disposition Costs..... 19
8.5 Disposal of Tangible Property..... 19

Article 9. LIABILITY..... 19

9.1 Extent of Government Liability..... 19
9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Liability..... 19
9.3 *Force Majeure*..... 20

Article 10. GENERAL PROVISIONS..... 20

10.1 Characteristics of the Agreement..... 20
10.1.1 Entire Agreement..... 20
10.1.2 Severability..... 20
10.1.3 Interpretation of Headings..... 20
10.2 Agreement Between Collaborators..... 20
10.2.1 Governing Laws..... 20
10.2.2 Independent Parties/Entities..... 20
10.2.3 Assignment /Subcontracting..... 20

10.2.4 Disputes..... 21
10.2.4.1 Settlement and Resolution..... 21
10.2.4.2 Continuation of Cooperative Work..... 21
10.2.5 Waivers..... 21
10.2.6 Use of Name or Endorsements..... 22
10.3 Environment, Safety, and Health..... 22
10.4 U.S. Competitiveness..... 22
10.5 Public Release of This Agreement..... 22

Article 11. MODIFICATIONS AND NOTICES..... 22

11.1 Amendments..... 22
11.2 Termination..... 23
11.2.1 Termination by Mutual Consent..... 23
11.2.2 Unilateral Termination..... 23
11.3 Notices..... 23

Article 12. SURVIVING PROVISIONS..... 24

Article 13. DURATION..... 24

Article 14. SIGNATURES..... 24

Appendix A
Statement of Work

Appendix B
Confirmatory License Agreement

STANDARD
NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY
RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS
(MURI / OPUS) LABORATORY

PREAMBLE

Under authority of the U.S. Federal Technology Transfer Act of 1986 (Public Law 99-502, 20 October 1986, as amended), **NAVAIR ORLANDO TSD (Naval Air Warfare Center Training Systems Division)**, located at **12350 Research Parkway, Orlando, Florida 32826-3275**, and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, whose administrative offices are located at **12443 Research Parkway, Suite 207, Orlando, FL 32826-3252**, enter into this Cooperative Research and Development Agreement (CRADA), which shall be binding upon the Collaborators and their assignees according to the clauses and conditions hereof and for the term and duration set forth.

The U.S. Federal Technology Transfer Act of 1986, as amended, provides for making the expertise, capabilities, and technologies of U.S. Federal laboratories accessible to other Federal agencies; units of State or local government; industrial organizations (including corporations, partnerships and limited partnerships, and industrial development organizations); public and private foundations; nonprofit organizations (including universities); or other persons in order to improve the economic, environmental, and social well-being of the United States by stimulating utilization of U.S. Federally funded technology developments and/or capabilities.

NAVAIR ORLANDO TSD has extensive expertise, capabilities, and information in **Virtual Environments for Operator Training**, and, in accordance with the U.S. Federal Technology Transfer Act, desires to make this expertise and technology available for use in the public and private sectors.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory has the interest, resources, capabilities, and technical expertise to transition the results of Naval research and development for public use.

NOW THEREFORE, the Collaborators agree as follows.

Article 1. DEFINITIONS

As used in this Agreement, the following terms shall have the meanings defined below, which are equally applicable to both the singular and plural forms of nouns or any tense of verbs.

1.1 "Agreement" means this Cooperative Research and Development Agreement (CRADA) with its Appendices.

1.2 "Classified Information" means all Data classified in accordance with the national security laws of the United States.

1.3 "Collaborator" means the Navy participant or the Non-Navy participant represented and bound by the signatories of this Agreement.

1.4 "Controlled Unclassified Information (CUI)" means Government Data, Information, or materials provided to or resulting from this Agreement that may be export controlled, sensitive, for official use only, or otherwise protected by law, executive order, or regulation.

1.5 "Cooperative Work" means research, development, engineering, or other tasks performed under this Agreement by **NAVAIR ORLANDO TSD** or **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** working individually or together, pursuant to the Objectives (Article 2) and the Statement of Work (Appendix A).

1.6 "Data" means recorded information of any kind regardless of the form or method of the recording, including computer software.

1.7 "Effective Date" means the date of the last signature of the Collaborators executing this Agreement.

1.8 "Exclusive License" means the grant by the owner of Intellectual Property of the exclusive right to make, use, or sell a patented invention.

1.9 "Government" means the Government of the United States of America.

1.10 "Government Purpose Rights" means the right of the Government to use, duplicate, or disclose Data, in whole or in part, and in any manner, for Government purposes only, and to have or permit others to do so for Government purposes only. Government Purpose Rights includes competitive procurement, but does not include the right to have or permit others to use Data for commercial purposes.

1.11 "Information" means all data, trade secrets, and commercial and financial information. (Chapter 5 Subsection II of Title 5 USC)

1.12 "Intellectual Property" means the property of ideas, examples of which include, but are not limited to, patents, trademarks, copyrights, and trade secrets.

1.13 "Invention" means any invention or discovery that is or may be patentable or otherwise protected under Title 35, United States Code, or any novel variety of plant that is or may be patentable under the Plant Variety Protection Act. (15 USC 3703(9))

1.14 "Invention Disclosure" means the document identifying and describing to organizational management the Making of an Invention.

1.15 "Made" when used in conjunction with any Invention means the conception or first actual reduction to practice of such Invention. (15 USC 3703(10))

1.16 "Militarily Critical Technologies (MCT)" means those technologies identified in the Militarily Critical Technologies List and under the Export Administration Act of 1979, as amended.

1.17 "Non-Subject Data" means any Data that are not Subject Data.

1.18 "Non-Subject Invention" means any Invention that is not a Subject Invention.

1.19 "Patent Application" means an application for patent protection for an Invention with any domestic or foreign patent-issuing authority.

1.20 "Principal Investigator (PI)" means that person having the responsibility for the performance of the Cooperative Work on behalf of a Collaborator.

1.21 "Proprietary Information" means information that embodies trade secrets developed at private expense or business, commercial, or financial information that is privileged or confidential provided that such information:

is not known or available from other sources without obligations concerning its confidentiality;

has not been made available by the owners to others without obligation concerning its confidentiality;

is not already available to the Government without obligation concerning its confidentiality; and

has not been developed independently by persons who have had no access to the information. (FAR/DFARS Definition)

1.22 "Restricted Access Information" means Subject Data generated by **NAVAIR ORLANDO TSD** that would be Proprietary Information if the Information had been obtained from a non-Federal Collaborator participating in a CRADA (15 USC 3710a). Under 15 USC 3710a(c)(7)(B), the Collaborators mutually may agree to provide appropriate protection to Subject Data generated by **NAVAIR ORLANDO TSD** (Restricted Access Information) against public dissemination or release under the Freedom of Information Act (FOIA) for a period of up to five (5) years after development of the Information.

1.23 "Subject Data" means that Data first recorded in the performance of the Cooperative Work.

1.24 "Subject Invention" means any Invention Made in the performance of the Cooperative Work.

1.25 "Tangible Property" means personal or real property that can be physically touched or held.

1.26 "Unlimited Rights" means the right to use, modify, reproduce, release, disclose, perform, or display Data or Computer Programs in whole or in part, in any manner and for any purpose whatsoever, and to have or permit others to do so.

Article 2. OBJECTIVES

Background

NAVAIR ORLANDO TSD has been conducting Navy sponsored research in virtual environments for operator/small arms training for over 15 years. Results of this research have already been transitioned for many military training applications, and have also been the basis for technology transition in use by law enforcement. The Small Arms Simulation Training (SAST) laboratory at **NAVAIR ORLANDO TSD** now has in use an extensive range of software and data bases that have resulted from in-house **NAVAIR ORLANDO TSD** research work.

The University of Central Florida (UCF), with plans to establish a new laboratory for Human Factors research related to Homeland Defense, became aware of the work already done in **NAVAIR ORLANDO TSD's** SAST laboratory. Discussions were immediately begun to explore the mutual benefits of an agreement to work collaboratively.

The primary objective of this work will be to enhance mutual understand of operator performance under stress and develop possible applications of human factors research in this area for military training, and for homeland defense and related technologies

Benefit to Navy Partner: Research using **NAVAIR ORLANDO TSD** developed or enhanced source code may be expected to result in further enhancements to software and databases, at no expense to the government, and with direct use in military simulation-based training. Additionally, the anticipated UCF human factors research projects may be expected to have direct application and benefit for military training. Further, expertise will be provided to UCF on a cost basis, leveraging Navy research dollars.

Benefit to Non-Navy partner: **NAVAIR ORLANDO TSD** developed or enhanced source code for scenario generation, rural and urban databases, weapon tracking, image generation, networking/communication, etc., can provide the kind of virtual environment test bed required for planned human factors research, and can greatly accelerate work toward research objectives.

Article 3. RESPONSIBILITIES

The Collaborators shall provide personnel, facilities, and equipment necessary for, and shall perform, the Cooperative Work.

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities

The Cooperative Work done by **NAVAIR ORLANDO TSD** will be performed under the program guidance of **Ron Wolff**, PI, **NAVAIR ORLANDO TSD, AIR 4962**, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed

within the facilities of **NAVAIR ORLANDO TSD** or done on behalf of **NAVAIR ORLANDO TSD** by third parties in support of this Agreement.

UCF MURI / OPUS Laboratory personnel who perform Cooperative Work at **NAVAIR ORLANDO TSD** facilities will be supervised by the **UCF MURI / OPUS Laboratory** PI.

3.2 **UCF MURI / OPUS Laboratory** Personnel and Facilities

The Cooperative Work done by the **UCF MURI / OPUS Laboratory** will be performed under the program guidance of **Dr. Peter Hancock, UCF MURI / OPUS Laboratory** PI, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed within the facilities of the **UCF MURI / OPUS Laboratory** or done on behalf of the **UCF MURI / OPUS Laboratory** by third parties in support of this Agreement.

NAVAIR ORLANDO TSD personnel who perform Cooperative Work at the **UCF MURI / OPUS Laboratory** facilities will be supervised by **NAVAIR ORLANDO TSD** PI.

3.3 Security Regulations and Directives

Each Collaborator will abide by the safety and security regulations and directives of the host facility in which the Cooperative Work is being performed.

Article 4. REPRESENTATIONS AND WARRANTIES

4.1 **NAVAIR ORLANDO TSD**'s Representations and Warranties

NAVAIR ORLANDO TSD hereby warrants and represents to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** as follows:

4.1.1 **NAVAIR ORLANDO TSD** is a Federal laboratory of the U.S. Department of the Navy (Navy) as defined by 15 USC 3710a(d)(2)(A) and Department of Defense Instruction 5535.8, dated May 14, 1999.

4.1.2 The performance of the activities specified by this Agreement is consistent with the **Virtual Environments for Operator/Small Arms Training** and technology transfer missions of **NAVAIR ORLANDO TSD** (15 USC 3710a).

4.1.3 The Department of the Navy official executing this Agreement for **NAVAIR ORLANDO TSD** has the requisite power and authority to enter into this Agreement and to bind **NAVAIR ORLANDO TSD** to perform according to the terms of this Agreement.

4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Representations and Warranties

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory hereby warrants and represents to **NAVAIR ORLANDO TSD** as follows:

4.2.1 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not directly or indirectly controlled by a foreign company or government (Executive Order 12591, Section 4 (a)). **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, as of the Effective Date of this Agreement, is a part of a **university** duly organized, validly existing, and in good standing under the laws of **the State of Florida.**

4.2.2 The official executing this Agreement for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** has the requisite power and authority to enter into this Agreement and to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to perform according to the terms of this Agreement.

4.2.3 The University of Central Florida and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** have taken all actions required by law, or otherwise, to authorize the execution and delivery of agreements, such as this Agreement.

4.2.4 The execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any agreement binding on **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. Furthermore, the execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any valid order of any court, or any regulatory agency or other body having authority to which **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is subject.

4.2.5 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not presently subject to debarment or suspension by any agency of the Government. Should **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** be debarred or suspended during the term of this Agreement or thereafter, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** will notify **NAVAIR ORLANDO TSD** within thirty (30) days of receipt of a final notice. **NAVAIR ORLANDO TSD** may then elect to terminate this Agreement and any licenses and options granted under this Agreement.

4.2.6 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is not** a small business as defined in 15 USC 632 and implementing regulations (13 CFR 121.101 et seq.) of the Administrator of the Small Business Administration.

4.3 Joint Representations

The Collaborators make the following representations.

STANDARD
NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY
RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS
(MURI / OPUS) LABORATORY

PREAMBLE

Under authority of the U.S. Federal Technology Transfer Act of 1986 (Public Law 99-502, 20 October 1986, as amended), **NAVAIR ORLANDO TSD (Naval Air Warfare Center Training Systems Division)**, located at **12350 Research Parkway, Orlando, Florida 32826-3275**, and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, whose administrative offices are located at **12443 Research Parkway, Suite 207, Orlando, FL 32826-3252**, enter into this Cooperative Research and Development Agreement (CRADA), which shall be binding upon the Collaborators and their assignees according to the clauses and conditions hereof and for the term and duration set forth.

The U.S. Federal Technology Transfer Act of 1986, as amended, provides for making the expertise, capabilities, and technologies of U.S. Federal laboratories accessible to other Federal agencies; units of State or local government; industrial organizations (including corporations, partnerships and limited partnerships, and industrial development organizations); public and private foundations; nonprofit organizations (including universities); or other persons in order to improve the economic, environmental, and social well-being of the United States by stimulating utilization of U.S. Federally funded technology developments and/or capabilities.

NAVAIR ORLANDO TSD has extensive expertise, capabilities, and information in **Virtual Environments for Operator Training**, and, in accordance with the U.S. Federal Technology Transfer Act, desires to make this expertise and technology available for use in the public and private sectors.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory has the interest, resources, capabilities, and technical expertise to transition the results of Naval research and development for public use.

NOW THEREFORE, the Collaborators agree as follows.

Article 1. DEFINITIONS

As used in this Agreement, the following terms shall have the meanings defined below, which are equally applicable to both the singular and plural forms of nouns or any tense of verbs.

1.1 "Agreement" means this Cooperative Research and Development Agreement (CRADA) with its Appendices.

1.2 "Classified Information" means all Data classified in accordance with the national security laws of the United States.

1.3 "Collaborator" means the Navy participant or the Non-Navy participant represented and bound by the signatories of this Agreement.

1.4 "Controlled Unclassified Information (CUI)" means Government Data, Information, or materials provided to or resulting from this Agreement that may be export controlled, sensitive, for official use only, or otherwise protected by law, executive order, or regulation.

1.5 "Cooperative Work" means research, development, engineering, or other tasks performed under this Agreement by **NAVAIR ORLANDO TSD** or **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** working individually or together, pursuant to the Objectives (Article 2) and the Statement of Work (Appendix A).

1.6 "Data" means recorded information of any kind regardless of the form or method of the recording, including computer software.

1.7 "Effective Date" means the date of the last signature of the Collaborators executing this Agreement.

1.8 "Exclusive License" means the grant by the owner of Intellectual Property of the exclusive right to make, use, or sell a patented invention.

1.9 "Government" means the Government of the United States of America.

1.10 "Government Purpose Rights" means the right of the Government to use, duplicate, or disclose Data, in whole or in part, and in any manner, for Government purposes only, and to have or permit others to do so for Government purposes only. Government Purpose Rights includes competitive procurement, but does not include the right to have or permit others to use Data for commercial purposes.

1.11 "Information" means all data, trade secrets, and commercial and financial information. (Chapter 5 Subsection II of Title 5 USC)

1.12 "Intellectual Property" means the property of ideas, examples of which include, but are not limited to, patents, trademarks, copyrights, and trade secrets.

1.13 "Invention" means any invention or discovery that is or may be patentable or otherwise protected under Title 35, United States Code, or any novel variety of plant that is or may be patentable under the Plant Variety Protection Act. (15 USC 3703(9))

1.14 "Invention Disclosure" means the document identifying and describing to organizational management the Making of an Invention.

1.15 "Made" when used in conjunction with any Invention means the conception or first actual reduction to practice of such Invention. (15 USC 3703(10))

1.16 "Militarily Critical Technologies (MCT)" means those technologies identified in the Militarily Critical Technologies List and under the Export Administration Act of 1979, as amended.

1.17 "Non-Subject Data" means any Data that are not Subject Data.

1.18 "Non-Subject Invention" means any Invention that is not a Subject Invention.

1.19 "Patent Application" means an application for patent protection for an Invention with any domestic or foreign patent-issuing authority.

1.20 "Principal Investigator (PI)" means that person having the responsibility for the performance of the Cooperative Work on behalf of a Collaborator.

1.21 "Proprietary Information" means information that embodies trade secrets developed at private expense or business, commercial, or financial information that is privileged or confidential provided that such information:

is not known or available from other sources without obligations concerning its confidentiality;

has not been made available by the owners to others without obligation concerning its confidentiality;

is not already available to the Government without obligation concerning its confidentiality; and

has not been developed independently by persons who have had no access to the information. (FAR/DFARS Definition)

1.22 "Restricted Access Information" means Subject Data generated by **NAVAIR ORLANDO TSD** that would be Proprietary Information if the Information had been obtained from a non-Federal Collaborator participating in a CRADA (15 USC 3710a). Under 15 USC 3710a(c)(7)(B), the Collaborators mutually may agree to provide appropriate protection to Subject Data generated by **NAVAIR ORLANDO TSD** (Restricted Access Information) against public dissemination or release under the Freedom of Information Act (FOIA) for a period of up to five (5) years after development of the Information.

1.23 "Subject Data" means that Data first recorded in the performance of the Cooperative Work.

1.24 "Subject Invention" means any Invention Made in the performance of the Cooperative Work.

1.25 "Tangible Property" means personal or real property that can be physically touched or held.

1.26 "Unlimited Rights" means the right to use, modify, reproduce, release, disclose, perform, or display Data or Computer Programs in whole or in part, in any manner and for any purpose whatsoever, and to have or permit others to do so.

Article 2. OBJECTIVES

Background

NAVAIR ORLANDO TSD has been conducting Navy sponsored research in virtual environments for operator/small arms training for over 15 years. Results of this research have already been transitioned for many military training applications, and have also been the basis for technology transition in use by law enforcement. The Small Arms Simulation Training (SAST) laboratory at **NAVAIR ORLANDO TSD** now has in use an extensive range of software and data bases that have resulted from in-house **NAVAIR ORLANDO TSD** research work.

The University of Central Florida (UCF), with plans to establish a new laboratory for Human Factors research related to Homeland Defense, became aware of the work already done in **NAVAIR ORLANDO TSD's** SAST laboratory. Discussions were immediately begun to explore the mutual benefits of an agreement to work collaboratively.

The primary objective of this work will be to enhance mutual understand of operator performance under stress and develop possible applications of human factors research in this area for military training, and for homeland defense and related technologies

Benefit to Navy Partner: Research using **NAVAIR ORLANDO TSD** developed or enhanced source code may be expected to result in further enhancements to software and databases, at no expense to the government, and with direct use in military simulation-based training. Additionally, the anticipated UCF human factors research projects may be expected to have direct application and benefit for military training. Further, expertise will be provided to UCF on a cost basis, leveraging Navy research dollars.

Benefit to Non-Navy partner: **NAVAIR ORLANDO TSD** developed or enhanced source code for scenario generation, rural and urban databases, weapon tracking, image generation, networking/communication, etc., can provide the kind of virtual environment test bed required for planned human factors research, and can greatly accelerate work toward research objectives.

Article 3. RESPONSIBILITIES

The Collaborators shall provide personnel, facilities, and equipment necessary for, and shall perform, the Cooperative Work.

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities

The Cooperative Work done by **NAVAIR ORLANDO TSD** will be performed under the program guidance of **Ron Wolff,** PI, **NAVAIR ORLANDO TSD, AIR 4962**, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed

within the facilities of **NAVAIR ORLANDO TSD** or done on behalf of **NAVAIR ORLANDO TSD** by third parties in support of this Agreement.

UCF MURI / OPUS Laboratory personnel who perform Cooperative Work at **NAVAIR ORLANDO TSD** facilities will be supervised by the **UCF MURI / OPUS Laboratory** PI.

3.2 **UCF MURI / OPUS Laboratory** Personnel and Facilities

The Cooperative Work done by the **UCF MURI / OPUS Laboratory** will be performed under the program guidance of **Dr. Peter Hancock, UCF MURI / OPUS Laboratory** PI, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed within the facilities of the **UCF MURI / OPUS Laboratory** or done on behalf of the **UCF MURI / OPUS Laboratory** by third parties in support of this Agreement.

NAVAIR ORLANDO TSD personnel who perform Cooperative Work at the **UCF MURI / OPUS Laboratory** facilities will be supervised by **NAVAIR ORLANDO TSD** PI.

3.3 Security Regulations and Directives

Each Collaborator will abide by the safety and security regulations and directives of the host facility in which the Cooperative Work is being performed.

Article 4. REPRESENTATIONS AND WARRANTIES

4.1 **NAVAIR ORLANDO TSD**'s Representations and Warranties

NAVAIR ORLANDO TSD hereby warrants and represents to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** as follows:

4.1.1 **NAVAIR ORLANDO TSD** is a Federal laboratory of the U.S. Department of the Navy (Navy) as defined by 15 USC 3710a(d)(2)(A) and Department of Defense Instruction 5535.8, dated May 14, 1999.

4.1.2 The performance of the activities specified by this Agreement is consistent with the **Virtual Environments for Operator/Small Arms Training** and technology transfer missions of **NAVAIR ORLANDO TSD** (15 USC 3710a).

4.1.3 The Department of the Navy official executing this Agreement for **NAVAIR ORLANDO TSD** has the requisite power and authority to enter into this Agreement and to bind **NAVAIR ORLANDO TSD** to perform according to the terms of this Agreement.

4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Representations and Warranties

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory hereby warrants and represents to **NAVAIR ORLANDO TSD** as follows:

4.2.1 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not directly or indirectly controlled by a foreign company or government (Executive Order 12591, Section 4 (a)). **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, as of the Effective Date of this Agreement, is a part of a **university** duly organized, validly existing, and in good standing under the laws of **the State of Florida.**

4.2.2 The official executing this Agreement for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** has the requisite power and authority to enter into this Agreement and to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to perform according to the terms of this Agreement.

4.2.3 The University of Central Florida and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** have taken all actions required by law, or otherwise, to authorize the execution and delivery of agreements, such as this Agreement.

4.2.4 The execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any agreement binding on **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. Furthermore, the execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any valid order of any court, or any regulatory agency or other body having authority to which **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is subject.

4.2.5 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not presently subject to debarment or suspension by any agency of the Government. Should **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** be debarred or suspended during the term of this Agreement or thereafter, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** will notify **NAVAIR ORLANDO TSD** within thirty (30) days of receipt of a final notice. **NAVAIR ORLANDO TSD** may then elect to terminate this Agreement and any licenses and options granted under this Agreement.

4.2.6 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is not** a small business as defined in 15 USC 632 and implementing regulations (13 CFR 121.101 et seq.) of the Administrator of the Small Business Administration.

4.3 Joint Representations

The Collaborators make the following representations.

4.3.1 There is no express or implied warranty as to any research, Invention, or product, whether tangible or intangible. In particular, the Collaborators make no express or implied warranty as to the merchantability or fitness for a particular purpose of any research, Invention, or product, whether tangible or intangible. Likewise, the Collaborators make no express or implied warranty as to any Cooperative Work, Subject Invention, Subject Data, or other product resulting from the Cooperative Work.

4.3.2 The use and dissemination of Information and materials exchanged under this Agreement will be in accordance with all U.S. laws and regulations, including those pertaining to national security and export control. Nothing in this Agreement shall be construed as a license to export Information or to permit any disclosure in violation of law, regulation, or Department of Defense policy. The exporting Collaborator is responsible for obtaining any export licenses that may be required by U.S. Federal law.

Article 5. FUNDING

5.1 Payment Schedule

Consistent with the information in Appendix A, **NAVAIR ORLANDO TSD** will provide source code/databases developed or enhanced by **NAVAIR ORLANDO TSD** to the **MURI / OPUS** lab at no cost to **UCF**. If, at a later date during the period of this CRADA, there is a mutually determined need for additional expertise from **NAVAIR ORLANDO TSD**, then the following will apply. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees to pay **NAVAIR ORLANDO TSD** on a cost basis, as mutually determined to be required, during the period of this CRADA, as specified in Appendix A.

Checks will be payable to:

NAVAIR ORLANDO TSD, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments

Each check and its cover correspondence shall refer to Navy CRADA number "NCRADA-**NAVAIR ORLANDO TSD-02-037**."

Checks will be mailed to:

NAVAIR ORLANDO TSD, 12350 Research Parkway (AIR 4.9T Nimmo), Orlando, Florida 32826-3275, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments.

5.2 Insufficient and Excess Funds

NAVAIR ORLANDO TSD may discontinue performance under this Agreement if the funds provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** for performance by **NAVAIR ORLANDO TSD** are insufficient or are not provided as specified in Article 5.1. In the event **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to tender the Government the required payment within fifteen (15) days after its respective due date, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall be in default under this Agreement for failure to make payments. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on**

behalf of the MURI / OPUS Laboratory is in default for this reason, NAVAIR ORLANDO TSD shall notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory.** If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** does not cure the default within fifteen (15) days of mailing date of notice, **NAVAIR ORLANDO TSD** may proceed to terminate the Agreement in accordance with Article 11.2.2, may cancel any option for an Exclusive License to a Subject Invention, and may terminate any Exclusive License granted pursuant to this Agreement.

Funds that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** paid under Article 5.1 and that **NAVAIR ORLANDO TSD** has not obligated or expended at the time of completion, expiration, or termination of this Agreement shall be returned to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** after **NAVAIR ORLANDO TSD's** submission of a final fiscal report to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory.**

5.3 No New Commitments

NAVAIR ORLANDO TSD shall make no new commitments concerning this Agreement after receipt of a written termination notice from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in accordance with Article 11.2 and shall, to the extent practicable, cancel all outstanding commitments by the termination date. Should such cancellation result in any costs incurred by **NAVAIR ORLANDO TSD, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that such costs shall be chargeable against any funding that it provided to **NAVAIR ORLANDO TSD.**

5.4 Accounting Records

NAVAIR ORLANDO TSD shall maintain current accounts, records, and other evidence supporting all its expenditures against funding provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement and shall retain such records for at least twelve (12) months after the completion, expiration, or termination of this Agreement. **NAVAIR ORLANDO TSD** shall provide **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** a financial report within four (4) months after completion, expiration, or termination of this Agreement.

Article 6. REPORTS AND PUBLICATIONS

6.1 Final Report

The Collaborators shall submit to each other a final report within four (4) months of the completion, termination, or expiration of this Agreement that includes the results obtained and a list of all Subject Inventions Made.

6.2 Agreement to Confer Prior to Publication or Public Disclosure

The Collaborators agree to confer and consult prior to any publication or public disclosure of Subject Data to ensure that no Proprietary Information, Restricted Access Information, Government Classified Information, CUI, or MCT Information is released and that patent rights are not compromised. Prior to any such publication or public disclosure of Subject Data, each Collaborator shall be offered a period not less than fifteen (15) days and not to exceed thirty (30) days, unless otherwise mutually agreed in writing by the Collaborators, to review any proposed abstract, publication, presentation, or other document for public disclosure that contains Subject Data. For the purposes of this Article, the term "disclosure" shall include, but not be limited to, submission of any manuscript for peer review prior to publication. It is the responsibility of the Collaborator intending to make public disclosure of Subject Data to notify the other Collaborator of such intent.

If a Collaborator objects to a proposed public disclosure, that Collaborator must so notify the other Collaborator within thirty (30) days of the date of notice of intent to disclose publicly. If no objection is received by the Collaborator intending to make public disclosure, concurrence is assumed. If a Collaborator objects on the grounds that patent rights may be compromised, a Patent Application must be filed by the responsible Collaborator within ninety (90) days of the date of notification of intent to make public disclosure, or by another date mutually agreed to by the Collaborators. If a Collaborator objects to the release of Information on the grounds that the Information is Proprietary Information, Restricted Access Information, or Information whose dissemination is restricted by U.S. security laws or regulations, the disclosure shall be postponed until the Information no longer meets the definitions of Proprietary Information, Restricted Access Information, or is no longer covered by U.S. security laws or regulations.

6.3 Classified Information

Any presentation that includes Subject Data that are Classified Information or otherwise restricted Data must have prior review and approval by **NAVAIR ORLANDO TSD** pursuant to the pertinent security laws, regulations, and directives.

Article 7. INTELLECTUAL PROPERTY

7.1 Data

7.1.1 General Provisions Applying to All Data

7.1.1.1 Ownership

Each Collaborator shall have title to all Data generated by that Collaborator.

7.1.1.2 No Implied License

Unless otherwise specifically provided, the Collaborators agree that the exchange of Data of any kind does not confer a license to any Invention claimed in any

patent or Patent Application or to the subject matter of any copyright, trademark/service mark, or other form of Intellectual Property protection.

7.1.1.3 Marking of Data

7.1.1.3.1 Data Provided With Less Than Unlimited Rights

Each Collaborator shall mark all Data that it provides with less than Unlimited Rights with a marking that clearly identifies the limited rights.

7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall place a proper proprietary marking on each medium used for recording Data that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** delivers to **NAVAIR ORLANDO TSD** under this Agreement that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** asserts is Proprietary Information. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall request in writing if it wishes Subject Data generated by **NAVAIR ORLANDO TSD** to be marked as Restricted Access Information. The Collaborators together shall confer to determine if such marking is appropriate, with reference to the Definitions of Article 1. If the Collaborators mutually agree to the marking then:

(a) For Non-Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory – NAVAIR ORLANDO TSD** MAY USE ONLY FOR PURPOSE OF CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(b) For Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** – GOVERNMENT HAS GOVERNMENT PURPOSE RIGHTS UNDER CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(c) For Data that are Restricted Access Information, the marking shall read:

"RESTRICTED ACCESS INFORMATION – PROTECT IN ACCORDANCE WITH CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**" UNTIL **SEPTEMBER 2005**".

7.1.1.3.3 Data That are Subject to 35 USC 205

NAVAIR ORLANDO TSD shall mark Data it provides under this Agreement that disclose one or more Inventions in which the Government owns or may own a right, title or interest, and that are subject to confidentiality under 35 USC 205. Such Data shall be marked:

"**NAVAIR ORLANDO TSD** DATA PROTECTED FROM RELEASE OR DISCLOSURE UNDER 35 USC 205."

7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted

Each Collaborator shall mark all Data that are Classified Information, CUI, MCT, or otherwise restricted by U.S. security or export control laws or regulations that it provides under this Agreement.

7.1.1.4 Protection of Data

Except for the rights granted in Article 7.1.2.2, Data shall be protected in accordance with the proper markings of its owner and as provided by, at a minimum, the requirements of 15 USC 3710a. Proprietary Information will be protected only if it is properly marked as such. Information provided in intangible form that is Proprietary Information must be designated Proprietary Information at the time it is delivered, followed within fifteen (15) days by a writing summarizing the exact Information to be protected. The Collaborator receiving Information in an intangible form that is designated as Proprietary Information shall be responsible for protecting the Information as Proprietary Information during the fifteen (15) day notification period. After the fifteen (15) day period, if no written summary has been received, the receiving Collaborator need not continue to protect the Information received in intangible form.

Restricted Access Information shall be protected from public dissemination for up to five (5) years, as mutually agreed.

Classified Information, CUI, MCT, or otherwise restricted Information shall be protected in accordance with the security laws of the United States.

7.1.1.5 Release of Data Under the Freedom of Information Act

Data in the possession of **NAVAIR ORLANDO TSD** that are not marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or Restricted Access Information must be released by **NAVAIR ORLANDO TSD** where such release is required pursuant to a request under the Freedom of Information Act (FOIA) (5 USC 552). **NAVAIR ORLANDO TSD** shall protect Data that are properly marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI**

/ **OPUS Laboratory** or Restricted Access Information from release under the FOIA for as long as the marked Data meet the definition of CUI, Proprietary Information or Restricted Access Information. Prior to release of any such Data, **NAVAIR ORLANDO TSD** shall promptly notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** of any request for Data of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** regardless of whether the requested Data are marked Proprietary Information or as required under the public records law.

7.1.2 Subject Data

7.1.2.1 Delivery of Requested Subject Data

Each Collaborator shall have the right to review and receive delivery of all Subject Data generated by the other Collaborator. Requested Subject Data shall be delivered to the requesting Collaborator within fifteen (15) days of the request.

7.1.2.2 Rights in Subject Data

Except as represented in Article 4.3.2, the Collaborators shall have Unlimited Rights in all Subject Data that are not Proprietary Information or Restricted Access Information. Notwithstanding 15 USC 3710a, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** grants Government Purpose Rights in any Subject Data furnished by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to **NAVAIR ORLANDO TSD** under this Agreement that are properly marked as Proprietary Information. The Government has Government Purpose Rights in Subject Data that are Restricted Access Information.

7.1.3 Rights in Non-Subject Data

The Collaborators shall have Unlimited Rights in any Non-Subject Data delivered under this Agreement that are not Proprietary Information.

NAVAIR ORLANDO TSD has a limited right to use, reproduce, and disclose only to Government employees for use in support of the Cooperative Work any Non-Subject Data that are properly marked as Proprietary Information and are provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement. Such Proprietary Information can be used only for the purpose of performing the Cooperative Work unless consent to other use or disclosure is obtained from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in writing.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall have a limited right to use, reproduce, or disclose Non-Subject Data that may describe one or more Inventions in which the Government owns or may own a right, title or interest, if such Non-Subject Data are provided by **NAVAIR ORLANDO TSD** under this Agreement. In accordance with 35 USC 205, such Non-Subject Data are to be held in confidence. Such Non-Subject Data shall be properly marked by **NAVAIR ORLANDO TSD** and the limited rights of **The UCF Office of Research, UCF on behalf of its**

Board of Trustees on behalf of the MURI / OPUS Laboratory shall be defined by a separate non-disclosure agreement.

7.2 Copyrights

7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory may copyright works of authorship prepared pursuant to this Agreement if eligible for copyright protection under Title 17 USC.

7.2.2 Copyright License to the Government

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory grants to the Government a nonexclusive, irrevocable, paid-up license in copyrighted works of authorship, including software (17 USC 106) prepared pursuant to this Agreement for any purpose, consistent with the rights in Data described in Article 7.1.

7.2.3 Copyright Statement

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall include the following statement on any text, drawing, mask work or other work of authorship, that may be copyrighted under 17 USC, that is created in the performance of this Agreement:

"The U.S. Government has a copyright license in this work pursuant to a Cooperative Research and Development Agreement with **NAVAIR ORLANDO TSD**."

7.3 Trademarks and Service Marks

7.3.1 Ownership of Trademarks and Service Marks

The Collaborator first establishing a trademark or service mark for goods or services with which the mark is used shall be considered the owner of the mark.

7.3.2 Obligation of Employees to Report Trademarks and Service Marks

Employees of both Collaborators shall report the adoption of a trademark or service mark associated with the Cooperative Work to their employer within thirty (30) days of the first use of the mark. Use includes internal use of any product or service of the Cooperative Work.

7.3.3 Obligation of Collaborators to Notify Each Other

Each Collaborator shall notify the other Collaborator within thirty (30) days of their employee's report of the first use of a trademark or service mark.

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

Each Collaborator shall be entitled to own the Subject Inventions of its employees. Each Collaborator shall cooperate with the other Collaborator to obtain inventor signatures on Patent Applications, assignments or other documents required to secure Intellectual Property protection. For any Invention Made jointly by employees of the Collaborators, each Collaborator shall have ownership of the Subject Invention in the form of an undivided interest.

7.4.4 Filing of Patent Applications

7.4.4.1 Filing of Patent Applications on Solely Made Inventions

Each Collaborator has primary responsibility for filing Patent Applications on the Subject Inventions of its employee(s).

Notwithstanding such primary responsibility, by mutual agreement, the Collaborators may identify which Collaborator shall file a Patent Application on any Subject Invention.

7.4.4.2 Filing of Patent Applications on Jointly Made Inventions

In the case of an Invention jointly Made by employees of both Collaborators, the Collaborators shall confer and agree as to which Collaborator will file any Patent Application. Officers of the non-filing Collaborator shall cooperate with the filing Collaborator to obtain signatures on documents that are needed to file a Patent Application.

7.4.4.3 Preserving Intellectual Property Rights

The Collaborator responsible for filing of a Patent Application on any Subject Invention shall file such Patent Application at least sixty (60) days prior to any bar date or one year from the date the Invention Disclosure was received, whichever comes first. If no Patent Application is filed within the specified time period, the other Collaborator may assume control of filing the Patent Application and take title to the Subject Invention on ten (10) days written notification. The Collaborator that relinquished the responsibility to file shall retain a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Subject Invention practiced throughout the world by or on its behalf.

7.4.4.4. Filing Deadlines

The Collaborator responsible for filing any Patent Application for a Subject Invention shall notify the other Collaborator of all filing deadlines for prosecution of any Patent Application and maintenance of any patents on the Subject Invention. Notwithstanding the primary responsibility defined in Article 7.4.4.1, sixty (60) days prior to any filing deadline, the Collaborators shall confer to determine if the filing Collaborator intends to respond to the filing deadline. The non-filing Collaborator will be permitted to take action if the filing Collaborator declines.

7.4.4.5 Copies and Inspection

7.4.4.5.1 Copies of Prosecution Papers

Each Collaborator filing a Patent Application on a Subject Invention shall provide the other Collaborator with a copy of any communication relating

to prosecution of said Patent Application within thirty (30) days of receipt of such communication.

7.4.4.5.2 Access to Patent Application File and Right to Make Copies

Upon written request, the filing Collaborator shall give the other Collaborator an Associate Power of Attorney, with authorization to access the Patent Application, make copies, and, in the event the filing Collaborator fails or declines to take action, do all that is necessary to secure Intellectual Property protection for the Subject Invention.

7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application

In the event both Collaborators decline to file a Patent Application on a Subject Invention, the Government will renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

In the event both Collaborators decline to file a Patent Application on a Subject Invention, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** may, at its sole discretion, renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

7.4.5 Nonexclusive License to Subject Inventions

7.4.5.1 Nonexclusive License Grant

Each Collaborator grants to the other Collaborator a nonexclusive, irrevocable, paid-up license to practice a Subject Invention Made by employees of the granting Collaborator or have the Subject Invention practiced throughout the world by or on behalf of the other Collaborator. No nonexclusive license granted under this Agreement shall permit licensee to grant sublicenses.

7.4.5.2 Confirmatory Nonexclusive License Agreement

Each Collaborator has the obligation to provide a Confirmatory License Agreement (Appendix B) to the other Collaborator for each nonexclusive license within ninety (90) days of the date of filing.

7.4.6 Option for Exclusive License to Subject Inventions

NAVAIR ORLANDO TSD gives **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** the option of acquiring an Exclusive License for the field of use **(virtual environments for operator/small arms training)** in the Government's rights in any Subject Invention Made in whole or in part by a **NAVAIR ORLANDO TSD** employee. The license shall be for reasonable consideration. In

order to exercise this option, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must notify **NAVAIR ORLANDO TSD** in writing within one hundred and eighty (180) days of the filing of a Patent Application. Unless another time period is mutually agreed upon between the Collaborators, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must execute an Exclusive License to the Subject Invention within one hundred and eighty (180) days of election to exercise the option, or the Invention shall be made available for licensing by the public in accordance with 37 CFR Part 404.

Any Exclusive License granted by the Government in a Subject Invention is subject to the statutorily required reservation by the Government of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the Government (15 USC 3710a).

7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement

No license granted under this Agreement shall be assigned, licensed or otherwise disposed of except to the successor in interest of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which such license pertains.

7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions

7.4.8.1 Exclusive Licenses and Exclusive License Option

NAVAIR ORLANDO TSD may terminate any Exclusive License or cancel any option for an Exclusive License to a Subject Invention granted under this Agreement in the event that:

(a) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is in default for failure to make payment as agreed in Article 5; or

(b) The Agreement is terminated unilaterally by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**; or

(c) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to perform according to the Statement of Work (Appendix A); or

(d) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** becomes a foreign owned, controlled, or influenced (FOCI) organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.4.8.2 Nonexclusive Licenses

NAVAIR ORLANDO TSD shall terminate any nonexclusive license to a Subject Invention granted under this Agreement if **The UCF Office of Research,**

UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory becomes a FOCI organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.5 Non-Subject Inventions

7.5.1 Ownership of Non-Subject Inventions

Each Collaborator owns its Non-Subject Inventions.

7.5.2 Rights Under Other Agreements

Nothing in this Agreement is intended to change the rights in Intellectual Property acquired by the Collaborators in any other contract or agreement between the **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** and the Government.

7.5.3 No License to Non-Subject Inventions

This Agreement does not grant any Collaborator a license, express or implied, to any Non-Subject Invention.

7.6 Research License

Each Collaborator shall allow the other Collaborator to practice any of its Non-Subject Inventions for the purpose of performing the Cooperative Work.

No license, express or implied, for commercial application(s) is granted to either Collaborator in Non-Subject Inventions by performing the Cooperative Work.

For commercial application(s) of Non-Subject Inventions, a license must be obtained from the owner.

Article 8. TANGIBLE PROPERTY

8.1 Title to Preexisting Tangible Property

Each Collaborator shall retain title to all Tangible Property to which it had title prior to the Effective Date of this Agreement.

8.2 Tangible Property Purchased by Collaborators to Perform the Cooperative Work

Each Collaborator shall retain title to all Tangible Property that it purchases during the period of this Agreement. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** cannot take title to any Government Tangible Property under this Agreement. Collaborator consumables to be used in the Cooperative Work of this Agreement are the property of the purchasing Collaborator until consumed.

8.3 Title to Developed Tangible Property

All Tangible Property developed under this Agreement with all components purchased by one Collaborator shall be the property of that Collaborator. Tangible Property having any component purchased by **NAVAIR ORLANDO TSD** shall be the property of the Government, unless such Tangible Property can reasonably be separated without damage to the other individual components. After this Agreement is completed, expired, or terminated, if separation of components can be made without damage, the Collaborators may, by mutual agreement, separate the Tangible Property into its components and the separated components shall remain the property of the Collaborator that purchased them.

8.4 Tangible Property Operational and Disposition Costs

During the period of and upon completion, expiration, or termination of this Agreement, each Collaborator shall be responsible for all costs of maintenance, removal, storage, repair, disposal, and shipping of all Tangible Property to which it has title.

8.5 Disposal of Tangible Property

Disposal of Tangible Property shall be in accordance with applicable U.S. Federal, State, and local property disposal laws, environmental laws, and regulations.

Article 9. LIABILITY

9.1 Extent of Government Liability

The Government shall be solely liable for the negligent or wrongful acts of its officers and employees to the extent provided for in the Federal Tort Claims Act (28 USC 2671 et. seq.) and in other applicable laws and regulations of the United States that specifically waive sovereign immunity. Nothing in this Agreement shall be construed as a waiver of the sovereign immunity of the United States.

9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Liability

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is solely responsible for its actions and the actions of those acting for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in the performance of this Agreement and for any damages that may arise from any suit, action, or claim, and for any costs from or incidental to any suit, action, or claim, including but not limited to settlement and defense costs. Further, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that in any suit, action or claim brought by anyone not a party to this Agreement based on actions of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not pursue any actions to enter the Government as a party in such suit, action or claim unless the Government has some liability under the Federal Tort Claims Act.

9.3 *Force Majeure*

No Collaborator shall be liable for the consequences of any *force majeure* that (1) is beyond its reasonable control; (2) is not caused by the fault or negligence of such Collaborator; (3) causes such Collaborator to be unable to perform its obligations under this Agreement; and (4) cannot be overcome by the exercise of due diligence. In the event of the occurrence of a *force majeure*, the Collaborator unable to perform shall promptly notify the other Collaborator. The Collaborators shall suspend performance only for such period of time as is necessary to overcome the result(s) of the *force majeure* and shall use their best efforts to resume performance as quickly as possible.

Article 10. GENERAL PROVISIONS

10.1 Characteristics of the Agreement

10.1.1 Entire Agreement

This Agreement constitutes the entire agreement between the Collaborators concerning the Cooperative Work and supersedes any prior understanding or written or oral agreement relative to the Cooperative Work.

10.1.2 Severability

The illegality or invalidity of any Article of this Agreement shall not impair, affect, or invalidate any other Article of this Agreement.

10.1.3 Interpretation of Headings

Headings of the Articles of this Agreement are for convenience of reference only and do not form a part of this Agreement and shall in no way affect the interpretation thereof.

10.2 Agreements Between Collaborators

10.2.1 Governing Laws

United States Federal Laws shall govern this Agreement for all purposes.

10.2.2 Independent Parties/Entities

The relationship of the Collaborators to this Agreement is that of independent parties and not as agents of each other, partners, or participants in a joint venture. Each Collaborator shall maintain sole and exclusive control over its personnel and operations.

10.2.3 Assignment/Subcontracting

10.2.3.1 Neither Collaborator may allow third parties to perform any part of the Cooperative Work under this Agreement without express written consent of the other

Collaborator. If consent is obtained, the Collaborator requesting such consent shall remain fully responsible for the portion of the Cooperative Work to be accomplished under a third-party agreement, and the third party is not a Collaborator of this Agreement. Any third-party agreement to perform a portion of the Cooperative Work shall contain terms consistent with this Agreement.

10.2.3.2 This Agreement shall not be assigned or otherwise transferred by either Collaborator without the prior written consent of the other Collaborator, except to the successor of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which this Agreement pertains.

10.2.3.3 If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee is a U.S. company, and becomes, during the term of this Agreement or thereafter, directly or indirectly owned, controlled, or influenced by a foreign company or government (FOCI), then **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee shall promptly notify **NAVAIR ORLANDO TSD** to that effect.

10.2.4 Disputes

10.2.4.1 Settlement and Resolution

NAVAIR ORLANDO TSD and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agree to use reasonable efforts to reach a fair settlement of any dispute. If such efforts are unsuccessful, remaining issues in dispute will be referred to the signatories or their successors for resolution. If a dispute continues, the remaining issues may be submitted to the Chief of Naval Research (CNR), or the CNR designee, for resolution. This Agreement does not prevent any Collaborator from pursuing disputes in a U.S. Federal court of competent jurisdiction. No Collaborator will pursue litigation in a U.S. Federal court until after the CNR, or the CNR designee, decides the dispute, or until sixty (60) days after the dispute was first submitted to the CNR, or the CNR designee, whichever comes first.

10.2.4.2 Continuation of Cooperative Work

If payments or installment payments are to be made as stated under Article 5, **NAVAIR ORLANDO TSD** will not start or continue cooperative work until payments or installment payments are received.

10.2.5 Waivers

None of the provisions of this Agreement shall be considered waived by either Collaborator unless such waiver is given in writing to the other Collaborator, signed by the executing official of this Agreement or the official's successor having the authority to bind the Collaborator making the waiver. The failure of either Collaborator to insist upon strict performance of any of the terms and conditions herein, or failure or delay to exercise any rights provided herein or by law shall not be deemed a waiver of any right of either Collaborator under this Agreement.

10.2.6 Use of Name or Endorsements

Except as provided for in Article 7.2.3, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not use the name of **NAVAIR ORLANDO TSD** or any other Government entity on any product or service that is directly or indirectly related to either this Agreement or any patent license or assignment associated with this Agreement without the prior approval of **NAVAIR ORLANDO TSD**. By entering into this Agreement, **NAVAIR ORLANDO TSD** does not directly or indirectly endorse any product or service provided, or to be provided, by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, its successors, assignees, or licensees. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not in any way imply that the Department of the Navy endorses any such product or service.

10.3 Environment, Safety, and Health

Each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste in its custody during the course of this Agreement. At the conclusion of this Agreement, each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste still in its possession. Each Collaborator shall obtain at its own expense all necessary permits and licenses as required by U.S. Federal, State, and local law and shall conduct such handling, control, and disposition in a lawful and environmentally responsible manner. Each Collaborator is responsible for all required environmental, safety, and health compliance, notice, and monitoring related to its facility in accordance with U.S. Federal, State, and local law and regulations. Collaborators shall abide by the environmental, safety, and health directives of the host facility in which the Cooperative Work is being performed, and any U.S. Federal, State, or local laws and regulations pertaining to environment, safety, and health that are applicable to the host facility.

10.4 U.S. Competitiveness

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory agrees that any product, process, or service using Intellectual Property arising from the performance of this Agreement shall be manufactured substantially in the United States.

10.5 Public Release of This Agreement

This Agreement, without funding information (Article 5) and Appendices, may be released to the public.

Article 11. MODIFICATIONS AND NOTICES

11.1 Amendments

If a Collaborator wishes to modify this Agreement, the Collaborators shall confer in good faith to determine the desirability of such modification. Such modification shall not be effective until a written amendment is signed by both executing officials of this Agreement or their successors.

11.2 Termination

11.2.1 Termination by Mutual Consent

The Collaborators may elect to terminate this Agreement at any time by mutual consent. Such termination shall not be effective until a written termination agreement is signed by both executing officials of this Agreement or their successors.

11.2.2 Unilateral Termination

A Collaborator may unilaterally terminate this entire Agreement at any time by giving the other Collaborator written notice signed by the executing official of this Agreement or his/her successor, not less than thirty (30) days prior to the desired termination date. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** unilaterally terminates this Agreement, any option for an Exclusive License to a Subject Invention and any Exclusive License to a Subject Invention granted by or pursuant to this Agreement shall simultaneously be terminated.

11.3 Notices

All notices pertaining to or required by Articles of this Agreement, except those pertaining solely to the prosecution of any patent, trademark, or service mark, shall be in writing and shall be signed by an authorized representative of the Technology Transfer Office for **NAVAIR ORLANDO TSD** or the preferred contact for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, and all such notices shall be delivered by hand, sent by courier with proper registration, or sent by certified mail, return receipt requested, with postage prepaid, addressed as follows:

If to **NAVAIR ORLANDO TSD**:

Ms. Patricia R. Nimmo
Technology Transfer Project Manager
NAVAIR ORLANDO TSD (AIR 49T)
12350 Research Parkway
Orlando, FL 32826-3275

If to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

Ms. Kim Smith
Senior Contract Specialist
Office of Research
University of Central Florida
Orlando Tech Center
12443 Research Parkway, Suite 207
Orlando, Fl 32826-3252

A Collaborator shall notify the other Collaborator of a change of address in the manner set forth above.

Notices pertaining solely to the prosecution of any patent, trademark, or service mark related to this Agreement shall be in writing and shall be signed by and sent to the Collaborator's legal counsel for Intellectual Property. Legal counsel for Intellectual Property for each Collaborator shall send a copy of any such notice to the Technology Transfer Office for he **NAVAIR ORLANDO TSD**. If either Collaborator fails to identify such counsel upon request, then such notices shall be sent to the points of contact specified above.

Article 12. SURVIVING PROVISIONS

The Articles covering Definitions, Representations and Warranties, Funding, Reports and Publications, Intellectual Property, Tangible Property, Liability, General Provisions, Modifications and Notices, and Surviving Provisions shall survive the completion, termination, or expiration of this Agreement.

Article 13. DURATION

This Agreement expires three years after its Effective Date, unless otherwise extended in writing according to the provisions of Article 11.

Article 14. SIGNATURES

For **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

I, the undersigned, am duly authorized to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002

By: M.J. Soileau

Title: Vice President, Office of Research

For the Department of the Navy:

I, the undersigned, by 15 USC 3710a and Navy regulations, am duly authorized to bind the U.S. Navy to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002.

By:

R. A. MOHLER
Title: Captain, U.S. Navy
Commanding Officer

Navy Organization: NAVAIR ORLANDO, TRAINING SYSTEMS DIVISION

This Page Intentionally Left Blank.

APPENDIX A

STATEMENT OF WORK
BETWEEN
NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND
THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH, UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

NAVAIR ORLANDO TSD will:

1. Provide, for use in the UCF MURI/OPUS laboratory, source code/databases developed or enhanced by NAVAIR ORLANDO TSD as part of Navy sponsored research and development, e.g., Host Software, Image Generator Software, Weapon Tracking Software, Networking/Communications Software, Data Collection/Analysis Software, Scenario Builder Software, Rural and Urban Databases

2. Provide expertise, on a cost basis and as mutually determined to be required, on utilization of government provided software and databases for UCF MURI/OPUS Laboratory research

3. Provide, on a cost basis and as mutually determined to be required, expertise for further hardware/software/database development, testing, and training applications

4. Participate, on a cost basis, in mutually selected specific research projects conducted under the auspices of the UCF MURI/OPUS Laboratory

5. Provide access to NAVAIR ORLANDO TSD customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

3. Mutually develop processes separate from this agreement to allow exchange of funds, should that be mutually determined to be required during the period of this CRADA

NOTE 1: The parties agree that Article 9.2 Extent of Liability, will be subject to and interpreted in accordance with Florida Statutes, Section 768.28.

NOTE 2: FOREIGN ACCESS TO TECHNOLOGY

All work described in this SOW shall be subject to the following with regard to foreign access to technology.

The Collaborators agree that research findings and technology developments under this CRADA may constitute a significant enhancement to the national defense. Accordingly, access to important technology developments under this CRADA by foreign firms or institutions must be carefully controlled. These controls are in addition to, and are not intended to change or supercede, the provisions of the International Traffic in Arms Regulation (22 CFR pt. 121 et seq.) the DoD Industrial Security Regulation (DoD 5220.22-R) and the Department of Commerce Export Regulation (15 CFR pt.770 et seq.)

The non-Navy Collaborator shall provide timely notice to the government's designated Principal Investigator of any proposed transfer of research findings and technology developments under this CRADA to foreign firms or institutions. If the government determines that the transfer may have adverse consequences to the national security interests of the United States, the non-Navy Collaborator and the government shall jointly endeavor to find alternatives to the proposed transfer which obviate or mitigate potential adverse consequences of the transfer but which provide substantially equivalent benefits to the non-Navy Collaborator.

APPENDIX B

CONFIRMATORY LICENSE AGREEMENT	*1. APPLICATION FOR (Title of Invention)*
2. INVENTOR(S) AND AFFILIATION	
3. PATENT APPLICATION SERIAL NO.	*4. PATENT APPLICATION FILING DATE*
5. NAVY ACTIVITY (Name, address, point of contact)	*6. NON-NAVY ACTIVITY (Name, address, point of contact)*
7. CRADA AGREEMENT NO.	*8. DATE OF THIS AGREEMENT*

9. The Invention identified above is a "Subject Invention" under Article 7 Intellectual Property included with the CRADA identified in Box 7 between the Department of the Navy and Non-Navy Activity identified in Box 6.

This document is confirmatory of the nonexclusive, irrevocable, paid-up license to practice the identified Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the receiving party, and of all other rights acquired by the receiving party by the referenced clause.

This license is granted to

_____ *the Government*

(Select one)

_____ *Non-Navy Activity identified in Box 6*

under this CRADA in the identified Invention, Patent Application and any resulting patent.

The licensee is hereby granted an irrevocable power to inspect and make copies of the above-identified Patent Application.

_____ *ACTIVITY NAME OF LICENSOR*

_____ *SIGNATURE*

_____ *NAME (Typed or Printed)*

_____ *TITLE*

_____ *BUSINESS TELEPHONE*

This Page Intentionally Left Blank.

University of Central Florida

Multi-Disciplinary University Research Initiative

MURI-OPUS

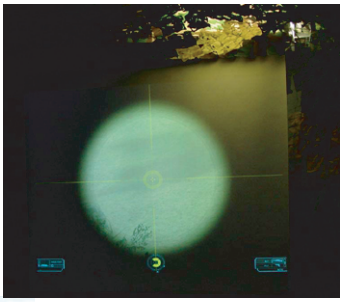
Operator Performance Under Stress

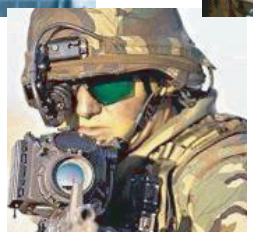
www.mit.ucf.edu

VOLUME FOUR











2004 Executive Summary

Table of Contents

NCRADA

A Collaborative Partnership between:

University of Central Florida

Contact Information:

University of Central Florida • MIT² Laboratory

3100 Technology Pkwy, Suite 337 • Partnership II • Orlando, Florida 32826-0544

http://www.mit.ucf.edu

Executive Summary

Accomplishments to Date

We are at the end of the third year of the Multiple University Research Initiative (*MURI*) program on Operator Performance Under Stress (*OPUS*). This report represents an account of progress to date, which contains numerous successes achieved at this stage. Following the original proposal, we have elaborated upon a number of areas that remain unresolved about the action of stress on performance. Each of these issues has served as a central theme in our on-going series of 'White Papers' that act to guide empirical research efforts. The three completed 'White Papers' are included in the present report and are to be followed by two additional 'White Papers' that, added to our initial model, will form the five pillars of our theoretical foundation. One of these white papers is near completion, and the last white paper is still in its early stages. Although we have detailed the progress of each discrete section of the MURI program individually, we continue to strengthen and exploit each of the theoretical, methodological, and electronic links between our respective laboratories. At the University of Central Florida, the MIT^2 research group has made considerable progress. We have been able to identify, generate, acquire and use our research test beds to evaluate our respective theoretical propositions. From the work done so far we now know much more about variations of spatio-temporal perception under stressful conditions. In addition, initial experiments testing Fuzzy Signal Detection Theory (FSDT) indicate that this novel approach to performance measurement captures elements of performance in complex tasks with high uncertainty and is a useful extension of traditional Signal Detection Theory. We have recently completed the mapping functions for a set of realistic-3d images of morphed images of tanks, those will be used to assess friend-foe identification in uncertain conditions. Our laboratory has also examined the role of individual differences in adaptation to stress associated with perceptual and cognitive tasks. Evidence to date indicates that personality traits of individuals influence their stress level and the coping strategies they employ to deal with the stress. We have completed a 'proof of concept' meta-analytic study of the effect of noise on performance on memory tasks. We are currently working on two additional meta-analytical studies of the effect of temperature and vibration on performance. These efforts are geared toward enhancing IMPRINT capacity for prediction of performance under stress. In addition, we have extended our field research to shooting performance at a police training shooting range. This facility allows us to measure stress in realistic environments similar to those of military soldiers in urban environments. We have recently completed a second field study in this facility. Data analysis is underway Through a Cooperative Research and Development Agreement (*CRADA*) with the Naval Air Warfare Center, Training systems Division, Orlando, the MIT^2 laboratory has completed installation of the Small Arms Simulation Test Bed (SAST) facility. We are currently using this facility for empirical investigation of cognitive changes that occur in a stressful target acquisition task. In addition, we are currently working with ARL HRED Laboratory on a possible collaborative simulation research which will upgrade the SAST and allow members of distributed teams to simultaneously participate in a common scenario. When linked together the systems of these two institutions can be used to examine the performance of larger teams. This initiative will allow the expansion of stress research to include social components that follow team performance under stressful situations.

Following our form of interaction, the chief member of the University of Minnesota group is now working at the University of Central Florida. The *VIKING* experimental environment developed at the University of Minnesota has been used to assess performance on tasks that

impose information processing demands similar to those faced by soldiers in combat environments. Currently the MIT^2 laboratory is using the recently developed ARES battery to assess the changes in cognitive performance of soldiers in the field during simulated combat, finding that many soldiers tolerate such field conditions with minimal impairment of cognitive performance if they know what is expected but that high demand events are still associated with cognitive performance changes. In addition, it appears that different forms of cognitive task are not affected equally in the context of combat. In addition, cases were observed in that study that clearly showed a precipitous decline in adaptive function at the extremes of stress, as predicted by the Hancock and Warm (1989) model. Recently, we have implemented the modified version of the ARES battery (ARES-II) that is more closely related to military tasks was integrated into our police shooting exercised, the results our currently being evaluated.

The Team Performance Laboratory (*TPL*) at the University of Central Florida has investigated adaptability of teams under time pressure and high workload, and is currently using a military-based task (*Flashpoint*) to further investigate the impact of stress on team processes. In particular, initial evidence indicates that the quality of team communication and performance varies as a function of team type (e.g., command and control versus a reconnaissance team). TPL researchers have also found that time pressure impairs communication among team members (breaking the closed-loop communication pattern).

MURI researchers at Linköping University (LiU), (formerly at Kansas State University) have investigated effects of trust on team decision-making, particularly, the identification and mitigation of cognitive decrements due to remote command and control under realistic simulations of battlefield conditions (on the paintball arena). These efforts have resulted in the identification of a viable research paradigm for the investigation of trust and team problem solving and represents one of the myriad examples of cross Laboratory technology transfer and protocol integration. The LiU group has conducted four simulated combat environment experiments on the effects of social interaction on performance. These experiments document a cognitive decrement directly attributable to remote command and control. They are currently conducting a new series of experiments that investigate a mechanism for mitigating the identified cognitive decrement. The success of the proposed mitigation has significant implications for the future of the dismounted infantry and for the design of remote command-and-control systems generally.

MURI researchers in the Cognitive Science Laboratory at the George Mason University (formerly at Catholic University of America) have completed their initial studies on the impact of automation on performance in decision-making tasks of the kind typically faced by military personnel in command and control settings. They found that as a result of complacency, automation with high (but not perfect) reliability does not always support performance. They also found that automation unreliability has a greater performance cost for decision automation than for information automation. In addition, they are exploring the effectiveness of adaptive automation using real-time measurement of workload under conditions of sustained workload and perceived stress. They are currently working on examining blood flow as an index of cognitive workload using the Transcranial Doppler Sonography System (TCDS). System sensitivity to blood flow changes in the left and right middle cerebral arteries is being examined in a working memory task involving storage and recall of digits and in the tracking sub-task of the Multi-Attribute Task Battery. Researchers at the University of Cincinnati have conducted

experiments demonstrating that the decline in sustained attention performance over time is accompanied by a parallel decline in cerebral bloodflow in the right hemisphere. In addition, they have initial evidence that transcranial oximetry (TCCO) measures of blood flow and oxygenation levels in the brain may also provide a useful physiological measure of sustained attention.

In respect of all these achievements, MURI has supported over thirty researchers, including three post-doctoral researchers and over twenty graduate student researchers across all participating Institutions.

Contributions to Scientific Understanding

We have made several fundamental contributions to the scientific understanding of stress effects. We have begun to elucidate the effects of diverse sources of real-world stress on spatio-temporal perception. Using stress sources that will impact further development of the IMPRINT model, we have begun to confirm our hypotheses on the 'narrowing' of the perceptual world of the stressed operator. These efforts have burgeoned into display design guidelines for such operations as a deliverable product, which has been published in *Ergonomics in Design*. We have made substantive theoretical progress on the question of adaptive interfaces for stressful operations. In particular, we have formulated a number critical questions about the efficacy of design and the possibility of optimal design. These developments have helped characterize and strengthen currently tested empirical propositions that promise to save significant dollars against a vision of endless, iterative search for design improvement. Our progress has been facilitated by naturalistic field observations and evaluations. The opportunity to conduct real-world test and evaluation on stressed soldiers and police officers accords with our theoretical proposition that a fully articulated theory can only emerge from an understanding of actual operations. While we continue to pursue laboratory and simulation efforts, we look to a union of forms of observation against which to test hypotheses formulated in the respective 'White Papers.' Our experiments on combined physical and cognitive effort have resulted in a surprising and important discovery. We call this the 'table-top' tolerance limit phenomenon. This phenomenon indicates that even when stress levels are significantly high the performance remains relatively constant to a point of sudden incipient failure. This finding has significant implications for both our future work and planning for military operations in general. Identifying symptomatic characteristics at the edge of failure (e.g., progressive increase in performance variability) may provide a predictive marker that enables the identification of failing operators. This would prove to be a crucial theoretical and practical discovery. We are also now beginning to distill ways in which team performance under stress is not simply the product of individual contributions. In particular, we are starting to show how workload sharing and task division and reallocation by team units acts to diffuse unacceptably high levels of demand. We are using this insight as a basis for designs in adaptive interfaces for stress in which demand may be shared between exposed soldiers and support personnel in remote locations. Initial developments in our collective work demonstrate how direct monitoring of operator state may be mapped into task resolution, well beyond the traditional conception of 'adaptive' interfaces. These new, neuro-ergonomic insights present the next generation level of conceptions for human-machine interaction. We have also found that stress influences mutual trust levels in combined decision-making conditions. The continuation of this effort seeks to explicate the specific mechanism of stress-induced deterioration. These represent only a selection of insights derived from the contained work, others of which are articulated in the specific papers.

Research for the Coming Year

In the coming year we look to complete the 'White Papers' as our theoretical efforts look toward reaching fruition. One of these papers is near completion, and the fifth White Paper will serve as a summary/integration of the previous four papers. Programmatic research evaluations related to the propositions set forth in the white papers will be pursued. Further experimentation is planned on the effects of time-pressure, cognitive workload change, task characteristics, and the degree of operator's control over the task environment on a variety of performance capacities. This experimentation permits tests of our theoretical models of the spatial and temporal components of stress and task performance, as well as the recruitment, allocation, depletion of resources and recovery under conditions of high workload and stress. Further, we plan further evaluation of target acquisition capacities and decision strategies as tests of fuzzy signal detection under stress. For this effort we have begun experimenting with basic shapes but also with more sophisticated and complex realistic objects and events. These studies will complement the psychophysical experiments already in progress and provide further evidence for the practical utility of FSDT. In addition, the FSDT model will be extended to include multidimensional signal detection theory concepts. To better understand the time-course of stress effects, future experimentation will examine the asymmetry in the rate of decline and recovery from performance under stress. Finally, the characteristics of operators that predict variation in performance will be evaluated in each of these studies. Such knowledge will inform development of selection and training procedures, as well as identify characteristics of individuals that aid in resisting the debilitative effects of stress. We are in the process of constructing a descriptive matrix that will inform current mechanisms for predicting task performance under stress (i.e., *IMPRINT*). Our efforts will evaluate stress-related performance deviations under vibration and heat conditions. Specifically, the ensuing database and meta-analytic review will aid in modifying and creating new degradation algorithms for current and future stressors within IMPRINT and mold the existing task taxonomy that is central to the effectiveness of performance shaping functions used within the task-network model. Our articulation of a complete theory and its instantiation in various modeling efforts is clearly contingent upon a much better understanding of the spectrum of influences experienced by actual soldiers. We shall continue our field efforts in order to understand and incorporate the nuances that the real-world presents, and continue to extend this work to include threat identification and shooting performance of police officers. Further, our field researchers help us identify those Forces personnel who would be the recipients and beneficiaries of our collective efforts.

Table of Contents

MURI Divisions

University of Central Florida (Peter Hancock)

Supported Work

Improving Seatbelts for Police and Military Personnel *Abstract* American Psychological Association Mid-Year Symposium Divisions 19 and 21 *Stafford, S.C., Oron-Gilad, T., Szalma, J.L., Delasontos, K., & Hancock, P.A. (2004)*

Effects of Spatial/Temporal Task Demands on False Alarms *Abstract* American Psychological Association Mid-Year Symposium Divisions 19 and 21 *Ross, J.M., Szalma, J.L., & Hancock, P.A. (2004)*

Effects of Spatial/Temporal Task Demands on False Alarms *Poster* American Psychological Association Mid-Year Symposium Divisions 19 and 21 *Ross, J.M., Szalma, J.L., & Hancock, P.A. (2004)*

Impact of Extraversion on Performance and Workload in Target Detection *Abstract* American Psychological Association Mid-Year Symposium Divisions 19 and 21 *Thropp, J.E., Szalma, J.L., Ross, J.M. & Hancock, P.A. (2004)*

Alternatives to Grouping Individuals in the Analysis of Individual Differences Data: Application to Optimism and Pessimism *Abstract* 110th Annual American Psychological Association Conference *Szalma, J.L., & Hancock, P.A. (2004)*

Alternatives to Grouping Individuals in the Analysis of Individual Differences Data: Applications to Dispositional Pessimism *Poster* 110th Annual American Psychological Association Conference *Szalma, J. L & Hancock, P. A. (2004)*

Attitudes of Police Officers Toward Seatbelt Use While on Duty *Abstract* 110th Annual American Psychological Association Conference *Stafford, S.C., Oron-Gilad, T., Szalma, J.L., Delasontos, K., & Hancock, P.A. (2004)*

What do we know About Teams? Toward a Practical Framework *Poster* 110th Annual American Psychological Association Conference *Priest, H. A., Burke, C. S., Salas, E. & Bowers, C. A. (2004)*

What do we know about Teams? A Practical Framework *Abstract* 110th Annual American Psychological Association Conference *Priest, H. A., Burke, C. S., Salas, E. & Bowers, C. A. (2004)*

Patterns of Team Communication: Are all Teams created Equal? *Presentation* 110th Annual American Psychological Association Conference *Priest, H. A., Guthrie, J. Burke, C. S., Bowers, C. A., Salas, E., & Milham, L. (2004)*

Patterns of Team Communication: Are all Teams created Equal? *Abstract* 110th Annual American Psychological Association Conference *Priest, H. A., Guthrie, J. Burke, C. S., Bowers, C. A., Salas, E., & Milham, L. (2004)*

Development of COTS Testbeds: Examining the Performance of Team Under Stress *Abstract* 110th Annual American Psychological Association Conference *Guthrie, Jr. J. W., Priest, H., Asberg, K., Burke, C. S., Salas, E. & Bowers, C. A. (2004)*

Development of COTS Testbeds: Examining the Performance of Team Under Stress *Presentation* 110th Annual American Psychological Association Conference *Guthrie, Jr. J. W., Priest, H., Asberg, K., Burke, C. S., Salas, E. & Bowers, C. A. (2004)*

The Influence of Dispositional Optimism and Pessimism on Task Engagement for Spatial and Temporal Discrimination *Abstract* 110th Annual American Psychological Association Conference *Szalma, J.L., Ross, J.M., & Hancock, P.A. (2004)*

The Influence of Dispositional Optimism and Pessimism on task engagement for spatial and Temporal Discrimination *Presentation* 110th Annual American Psychological Association Conference *Szalma, J. L., Ross, J. M., & Hancock, P. A. (2004)*

Analysis of Distributed Team Communication: Team Performance and Training Implications *Abstract* 110th Annual American Psychological Association Conference *Guthrie Jr., J.W., Asberg, K., & Bowers, C.A. (2004)*

Sequential Analysis of Distributed Team Communications: Implications for Team Performance and Training *Poster* 110th Annual American Psychological Association Conference *Guthrie, J. W., Asberg, K., Bowers, C. A., Salas, E. & Jentsch, F. (2004)*

Motivational Effects of Adding Context Relevant Stress in PC-Based Game Training *Refereed Publication* Morris, C. S., Hancock, P. A., & Shirkey, E. C. (2004) *Military Psychology*, **16** (2), 135-147.

Examination of Attentional Mechanisms Underlying Stress and Performance *Abstract* ARO Workshop-Symposium Cashiers *Szalma, J.L., Oron-Gilad, T. & Hancock, P.A. (2004)*

Protecting Soldiers: Warning Presentation and Retention Under Stress in a Military Task *Abstract* 24th Annual Army Science Conference *Helmick-Rich, J., Burke, K. A., Oron-Gilad, T., Moore, J., & Hancock, P. (2004)*

Protecting Soldiers: Warning Presentation and Retention Under Stress in a Military Task *Poster* 24th Annual Army Science Conference *Helmick-Rich, J., Burke, K. A., Oron-Gilad, T., Moore, J., & Hancock, P. (2004)*

Mapping an Objective State of the World to a Subjective Degree of Signalness: an Exercise in Fuzzy Signal Detection Theory Analysis *Presentation* 48th Annual Human Factors and Ergonomics Society *Morgan, J.F., Oron-Gilad, T., & Hancock, P.A. (2004)*

Are there differences in Teamwork across Team Type? *Proceedings* 48th Annual Human Factors and Ergonomics Society *Priest, H. A., Guthrie, J., Burke, C. S., Bowers, C., & Salas, E. (2004)*

Efficacy of Stress Exposure Training on Target Acquisition in Combat Simulations *Proceedings* 48th Annual Human Factors and Ergonomics Society *Ross, J.M., Ganey, H.C., & Broadway, R.S. (2004)*

Stress Effects on Soldier Performance *Panel* 48th Annual Human Factors and Ergonomics Society *Hancock, P.A. & Burke, S.-Organizers (2004)*

The Impact of Dispositional Optimism and Pessimism on Stress as a Function of Psychophysical Task Characteristics *Proceedings* 48th Annual Human Factors and Ergonomics Society *Ross, J.M., Szalma, J.L., & Hancock, P.A. (2004)*

Individual Differences Related to Shooting Performance and Perceived Workload in a Police Night-training Shooting Exercise *Proceedings* 48th Annual Human Factors and Ergonomics Society *Stafford, S.C., Oron-Gilad, T., Szalma, J.L., & Hancock, P.A. (2004)*

Workload and Performance: A Field Evaluation in a Police Shooting Range *Proceedings* 48th Annual Human Factors and Ergonomics Society *Greenwood-Erickson, A., Oron-Gilad, T., Szalma, J. L., Stafford, S. C., & Hancock, P. A. (2004)*

Performance Operating Characteristics for Spatial and Temporal Discriminations: Common or Separate Capacities? *Proceedings* 48th Annual Human Factors and Ergonomics Society *Thropp, J. E., Szalma, J. L., & Hancock, P. A. (2004)*

The Effects of Warning Presentation and Retention Under Varying Level of Stress *Proceedings* 48th Annual Human Factors and Ergonomics Society *Helmick-Rich, J. Burke, K. A., Oron-Gilad, T. & Hancock, P. (2004)*

Affect as an Aspect of Workload: Why and How to Address Affect in Automation and Simulation Studies *Paper* Human Performance, Situation Awareness and Automation Conference (HPSAA II) *Ganey, H.C.N., Koltko-Rivera, M. E., Murphy, L. L., Hancock, P.A., & Dalton, J. (2004)*

The Peacekeeper: How the Role of the Modern Soldier has Changed and How that Affects Workload *Paper* Human Performance, Situation Awareness and Automation Conference (HPSAA II) *Koltko-Rivera, M.E., Ganey, H.C.N., Murphy, L.M., Hancock, P.A., & Dalton, J. (2004)*

Menu Design Considerations for Rapid Decision-making Simulations *Presentation* Interservice/Industry Training, Simulation, and Education Conference *Pike, W., Burke, K.A., Helmick-Rich, J., Greenwood-Ericksen, A., & Bradley, K. (2004)*

Efficacy of Transfer in Simulation-Based Training: Implications for Stress Exposure Training *Paper* 2nd Swedish-American Workshop on Modeling and Simulation *Ross, J.M., Szalma, J.L., & Hancock, P.A. (2004)*

Fuzzy Signal Detection Theory: Basic and Applied Research *Poster* International Conference on Human Performance *Szalma, J. L., Stafford, S. C., Oron-Gilad, T., Murphy, L. L. & Hancock, P. A. (2004)*

Operator Performance Under Stress: A Multiple University Research *Poster* International Conference on Human Performance *Hancock, P. A., Szalma, J. L., Oron-Gilad, T., & the MURI-OPUS Group (2004)*

Fuzzy Signal Detection Theory: Tests of Assumptions and Comparison to Traditional Analyses *Abstract* Psychonomic Society Annual Conference *Murphy, L. L., Szalma, J. L., & Hancock, P. A. (2004)*

Fuzzy Signal Detection Theory: Tests of Assumptions and Comparison to Traditional Analyses *Poster* Psychonomic Society Annual Conference *Murphy, L. L., Szalma, J. L., & Hancock, P. A. (2004)*

Psychological Strategies for the Defence Against Terrorism. *Proceedings* NATO Symposium, Systems, Concepts and Integration Methods and Technologies for Defense Against Terrorism (RTO-MP-SCI-158). *Koltko-Rivera, M. E., & Hancock, P. A. (2004)*

Facilitated Work

Mental Rotation of Objects in an Airport Security Setting *Proceedings* 48th Annual Human Factors and Ergonomics Society *Morgan, J.F., Stafford, S., Nayeem, R., & Pepe, A. (2004)*

Personality Theory and Human Factors Research. *Proceedings* Human Performance, Situation Awareness and Automation Conference (HPSAA II). *Koltko-Rivera, M. E. (2004)*

Worldview and Acculturation as Predictors of Performance: Addressing These Variables in Human Factors / Ergonomics Research. *Proceedings* 48th Annual Human Factors and Ergonomics Society *Koltko-Rivera, M. E., Ganey, H. C. N., Dalton, J., & Hancock, P. A. (2004)*

On the Relevance of Qualitative Methods for Ergonomics. *Referred Publication* Hancock, P.A. & Szalma, J.L. (2004) *Theoretical Issues in Ergonomic Science, 5* (6), pp. 499-506.

Please Note: Wayne Harris is now with UCF.

Linköping Institute of Technology (Kip Smith)

Supported Work

The Adverse Impact of Remote Command and Control Under Fire *Refereed Publication (submitted)* The adverse impact of remote command and control under live fire. *Human Performance in Extreme Environments.*

Facilitated Work

Decision Support through Constraint Propagation in Collaborative Distributed Command and Control *Proceedings* 23rd European Annual Conference on Human Decision Making and Manual Control *Woltjer, R. & Smith, K. (2004).*

HUD use Improves Landing Performance in a Full-motion Simulator *Proceedings* 48th Annual Human Factors and Ergonomics Society *Goteman, Ö., Smith, K., & Dekker, S. (2004)*

Initial Experiments on Leader Presence and Communication Mode on Combat Performance *Proceedings* Human Performance, Situation Awareness, and Automation Technology, *Smith, K. (2004)*

Task Demand and Response Error in a Simulated Air Traffic Control Task: Implications For Ab Initio Training *Refereed Publication* Murphy, L. L., Smith, K., Hancock, P. A. (2004). Task demand and response error in a simulated air traffic control task: Implications for ab initio training. *International Journal of Applied Aviation Studies*, **4**(1), 91-106.

Managing risk in social exchange. *Book Chapter* Grazioli, S., Smith, K., & Johnson, P.E. (Eds.) *Psychological Explorations of Competent Decision Making* (pp. 71-123). Cambridge, UK: Cambridge University Press.

Psychological Explorations of Competent Decision Making *Edited Book* Smith, K., Shanteau, J., & Johnson, P.E. (Eds.) (2004) *Psychological Explorations of Competent Decision Making*. Cambridge, UK: Cambridge University Press.

Simulated touch: A Method for Stimulating Trust *Manuscript submitted for publication Smith, K., Liberg, M., Hilmersson, J. & Sitzes, R. (2004)*

The Value of Leader Presence *Manuscript submitted for publication Smith, K., & Pangburn, K. (2004)*

Time to Contact in Representational Space *Manuscript submitted for publication Smith, K. & Hancock, P. A. (2004)*

HUD Use Improves Landing Performance in Restricted Visibility. *Manuscript submitted for publication*. *Goteman, Ö., Smith, K., & Dekker, S. (2004)*

Progress Report Smith, K. 04-2004

Progress Report Smith, K. 10-2004

George Mason University (Raja Parasuraman)

Supported Work

The Effects of Event Rate and Time on Task on Human Performance and Subjective Ratings During a Multi-task Infantry Simulation. *Poster* Human Performance, Situation Awareness and Automation Conference (HPSAA II). *Zinni, M., Proto, R., Harris, W., & Parasuraman, R. (2004)*

Improving seatbelts for police and military personnel

S.C. Stafford, T. Oron-Gilad, J.L. Szalma, K. Delasontos, P.A. Hancock
University of Central Florida

The typical seatbelt design may interfere with operational work by lengthening response time to emergency situations. Thus, in high threat environments there may be a direct conflict between operational safety (effective responses to threat) and driver safety (seatbelt use). There may be situations where the usage of the seat belt does not enhance the officer's safety but creates a higher risk level by not enabling quick maneuvers and easy access to the weapon.

293 police officers from the southeastern U.S. filled a questionnaire that included work related and seatbelt usage information. For the current study, the effect of emergency and non-emergency situations on seatbelt usage was examined. A significant effect for emergency was found at short driving distances (emergency situations interfere with seatbelt usage).

These results support the hypothesis that a conflict between operational and driving safety exists. The data indicates that current seatbelt designs in police cruisers paradoxically represent a safety problem to police officers. Similarly, suitable seatbelts may be useful for other operational situations such as military vehicles or other domains that include high threat environments.

Effects of Spatial/Temporal Task Demands on False Alarms

J.M. Ross, J.L. Szalma, & P.A. Hancock
University of Central Florida

This study examined the relation between spatial and temporal dimensions in perception. Forty-six undergraduates performed three conjunction discrimination tasks. Two tasks required either a spatial or temporal discrimination. In the third task these dimensions were combined. Critical signals were cases in which both dimensions were at their critical value (e.g., short on both dimensions). In non-targets cases, one or both of the dimensions were presented at their 'neutral' level. Hit rate was not affected by spatial and temporal task properties, but there were significant effects for false alarm type. More false alarms were made when one dimension (spatial or temporal) was at its critical value than when both dimensions were at their non-target values. Further, significantly more false alarms were committed when the stimulus was presented at the critical value for the spatial dimension (and non-target value for the temporal dimension) than for the reverse condition. Thus, when paired together the spatial dimension dominates the temporal, possibly due to the greater salience of spatial information relative to temporal information. Future research will examine the degree of perceptual independence of these dimensions.

Effects of Spatial and Temporal Task Demands on False Alarms

J.M. Ross, J.L. Szalma, & P.A. Hancock

Abstract

This study examined the relation between spatial and temporal perceptual dimensions using three target detection tasks performed in the presence of white noise. Forty-six undergraduates performed three conjunction discrimination tasks. Two tasks required luminance discrimination in conjunction with either a spatial or temporal discrimination. In the third task both spatial and temporal discriminations were required. Targets for detection were cases in which both dimensions were at their target value (e.g., short on both dimensions). In non-targets cases, one or both of the dimensions were presented at their non-target level. Hit rate and overall false alarm rate were not affected by spatial and temporal task manipulation, but there were significant differences among false alarm type. More false alarms were made when one dimension (spatial or temporal) was at its target value than when both dimensions were at their non-target values. Further, significantly more false alarms were committed when the stimulus was presented at the target value for the spatial dimension (and non-target value for the temporal dimension) than for the reverse condition. Thus, when paired together the spatial dimension dominates the temporal, in terms of false alarms, possibly due to the greater salience of spatial information relative to temporal information. However, the temporal task induced greater perceived workload than the spatial, suggesting the effects of spatial and temporal manipulations are not equivalent. Future research will examine the degree of perceptual independence of these dimensions.

Introduction

In their model of stress and performance, Hancock and Warm (1989) proposed that individuals adapt to stress by the attentional narrowing mechanism articulated by Easterbrook (1959), in which the range of cue utilization is restricted in response to increased emotional arousal. They also argued that tasks themselves present the major source of stress, to which individuals adapt by the narrowing mechanism. They proposed two fundamental task dimensions that influence this adaptive process. These are information structure, often expressed spatially, and information rate, the temporal characteristics of the task. Recently, Hancock and Weaver (in press) have argued that the narrowing that occurs along the spatial dimension also occurs along the temporal, and that these represent a common attention mechanism. It was further argued by Hancock, Szalma, and Weaver (2002) that this results from perception (and distortions of those perceptions) of spatial and temporal information that draw on common resource capacities. Based upon these considerations, we tested the hypothesis that combining spatial and temporal demands within a discrimination task should debilitate performance and increase perceived workload as a result of this common resource mechanism, and that these effects should be exacerbated in the presence of an additional stressor (white noise). It was also hypothesized that if spatial and temporal perception share common resource capacities, combining these demands into a detection task should increase perceived workload relative to conditions in which each dimension is emphasized individually.

Method

Participants

Twenty-three females and twenty-three males, ranging in age from 18-36 (M = 23.3) at the University of Central Florida were paid for their participation in this study.

Experimental Design and Tasks

To test the aforementioned hypotheses two noise conditions (intermittent white noise vs. quiet) were factorially combined with three task types. The tasks required participants to discriminate between vertical lines varying along two of three dimensions: spatial (line length), temporal (duration of presentation of the line) and luminance of the line. The three tasks resulting from the pairwise combinations of dimensions consisted of 1) a task requiring a spatial discrimination and a luminance discrimination ('spatial-only' condition); 2) a task requiring a temporal discrimination and a luminance discrimination ('temporal only' condition); and 3) a task requiring both spatial and temporal discrimination (combined spatial-temporal condition). The luminance discrimination was used so that there would be an equal number of conjunctive searches across tasks. This experiment was a mixed design in which task type was a within subjects factor with three levels and noise was a between subjects factor with two levels. All participants experienced each of the three tasks and were assigned at random to one of the two noise and six possible task order conditions. Each task was seven minutes in duration.



Stimuli

The lines were presented on a background mask of .4mm diameter white circles on a black background (Temple, Warm, Dember, Jones, LaGrange, and Matthews, 2000). The tasks differed according to spatial and temporal properties of the stimuli. The spatial-only task required observers to detect a line of higher luminance *and* shorter length. Non-targets were presentations of a line at the higher luminance and of longer length, or a line of lower luminance and either length (see Figure 1). The duration of each stimulus was held constant at 200 milliseconds. The 'temporal-only' task required discrimination of both stimulus duration and luminance. Targets consisted of lines of higher luminance *and* shorter duration. Non-targets were lines of longer duration and lower luminance or lines of shorter duration at either luminance level. The length of the line stimuli in this condition was held constant at 3.5cm (see Figure 2). In the third task ('combined') the temporal discrimination was combined with spatial discrimination. Thus, observers were required to detect a line of the shorter duration *and* shorter length. The luminance of the stimuli was held constant.

The stimulus presentation rate was twenty-four events per minute with six signals per minute (signal probability=25%). The order of stimulus presentation was randomized for each participant. For each task participants responded to a signal by pressing the spacebar key on the keyboard. Responses occurring within 2.5 seconds of the onset of a signal were recorded as Correct Detection while all other responses were recorded as False Alarms.

Noise

For the noise group an 85 dBA intermittent white noise was presented via headphones during the three seven minute blocks. The intermittent white noise was composed of ten 4-second, eight 6-second, five 8-second, and four 10-second durations. The order and times of administration of noise was randomized across each session. The total duration of white noise per task was 2.8 minutes. Observers in the quiet condition also wore headphones to control for the effects of additional stimulation. The ambient sound level in the quiet condition was 39.5 dBA.

Perceived Workload

Perceived Workload was assessed using the NASA-Task Load Index (TLX; Hart & Staveland, 1988), a well regarded instrument that provides an index of overall workload as well as ratings reflecting the relative contributions of six sources of workload: Mental Demand, Temporal Demand, Physical Demand, Performance, Effort, and Frustration.

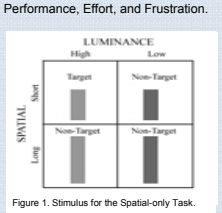


Figure 1. Stimulus for the Spatial-only Task.

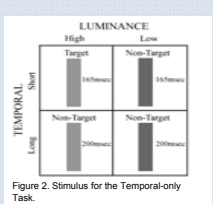


Figure 2. Stimulus for the Temporal-only Task.

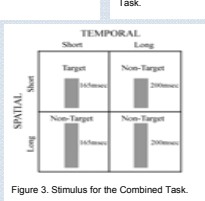


Figure 3. Stimulus for the Combined Task.

Results

Performance.

An ANOVA of the proportion of correct detections indicated no significant differences for task, noise, or the interaction between these variables (p>.05 in each case). Although an ANOVA of the proportion of false alarms indicated no significant effects for task, noise, or their interaction (p>.05), separate ANOVAs for false alarm type (see Figures 1-3) were computed for each task. In each case, *post-hoc* tests using the Bonferroni correction (α=.05) was used.

Spatial-Only Task

The effect for false alarm type was significant, $F(2,90)=22.07$, $p<.005$. *Post-hoc* tests indicated a higher false alarm rate (FAR) when either the spatial (FAR=0.31) or luminance (FAR=0.48) dimension was at its target value, compared to when the stimulus presented was at the non-target value for both dimensions (FAR=0.28). In addition, the FAR was higher when the spatial dimension was at its target value than when the luminance level was at its target value (see Figure 4).

Temporal-Only Task

An ANOVA for the 'Temporal-only' task also showed a main effect for false alarm type, $F(2,90)=5.47$, $p<.05$. *Post hoc* tests indicated that relative to the case in which both dimensions were at their non-target value (M=0.36), there was a higher propensity for false alarms when either the temporal (M=0.42) or luminance (M=0.44) dimension was at its target value. The false alarm rates for the latter two categories did not differ significantly from one another (p>.05; see Figure 5).

Combined Spatial-Temporal Task

An ANOVA for the Combined task indicated a significant effect for false alarm type, $F(2,90)=28.24$, $p<.005$. *Post-hoc* tests indicated a higher propensity for false alarms when the stimulus presented was spatial target/temporal non-target (M=0.45) or temporal target/spatial non-target (M=0.33), compared to when both dimensions were at their non-target value (M=0.26). In addition, the FAR associated with the spatial target/temporal non-target stimulus was significantly higher than that of the temporal target/spatial non-target stimulus (see Figure 6).

Perceived Workload.

An ANOVA of the Global workload ratings revealed a significant main effect for task, $F(2, 90) = 4.63$, $p < .05$. *Post-hoc* comparisons indicated that the Temporal-only task induced significantly more global workload than the combined task. In addition to global workload, weighted ratings on the subscales of the TLX were obtained. An ANOVA revealed a significant effect for task, $F(2, 90) = 3.91$, $p<.05$, with the 'Temporal-only' (M=208.17) task inducing significantly more workload as compared to the 'Combined task' (M=190.18). Significant effects were also observed for scales, $F(4,180)=15.18$, $p < .005$, and a task by scales interaction, $F(8,360)=2.05$, $p<.05$. Separate task by noise ANOVAS indicated significant task effects for the Weighted Temporal Demand subscale, $F(2,90)=6.55$, $p<.01$, and *post-hoc* comparisons indicated that the 'Spatial-only' task (M=129.67) induced significantly less temporal demand than either the 'Temporal-only' (M=195.22) or 'Combined' tasks (M=174.13), which did not differ significantly from one another.

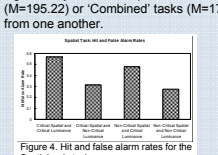


Figure 4. Hit and false alarm rates for the Spatial-only task.

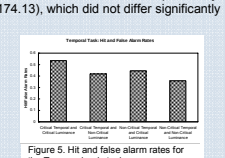


Figure 5. Hit and false alarm rates for the Temporal-only task.

Discussion

- Although the hypothesized differences were not observed in regard to performance or workload, there were unexpected differences among false alarm types in the spatial and combined conditions. The pattern of results suggests that the spatial-only task dominated attention over the temporal, since more errors were made when the line length was at its target value than when the non-target consisted of a time duration target value.

- Although no significant differences were observed in regard to overall performance, the temporal-only task induced more global workload than the combined task. Thus, replacing luminance discrimination with a spatial discrimination reduces global workload but does not modify overall performance. This represents a performance insensitivity (Hancock, 1996) and indicates that for observers in the temporal task performance was maintained at levels similar to those in the combined task, but only at the cost of greater resource investment. These results indicate that adding a spatial demand to a temporal demand does not draw significantly more capacity beyond that of the temporal task itself; indeed, inclusion of the spatial task *reduced* the workload.

- At the subscale level, *Weighted Temporal Demand* was greater in the temporal-only and combined tasks than in the spatial-only task, indicating greater temporal demand for a temporal discrimination task, regardless of whether a spatial discrimination is also required. These findings indicate that adding a temporal demand to a spatial discrimination increases workload.

- Taken together, the performance and workload effects suggest that spatial and temporal manipulations do not necessarily draw on common resource capacities. However, it is possible that distortion of spatial and temporal perception share a common narrowing mechanism but that their perceptual effects are not additive as hypothesized.

- In this experiment the white noise condition did not significantly impact performance. In a previous study on the effects of noise on performance of tasks varying in their spatial-temporal emphases, noise was found to increase leniency in responding (Ross, Szalma, Thropp, & Hancock, 2003). In this experiment no significant differences in hit rate or overall false alarm rate were observed, indicating that the noise effect reported by Ross et al. (2003) may depend on the particular manipulation of spatial and temporal properties. Note, however, that the tasks in this study were much more difficult than those employed by Ross et al. (2003). It may be that noise effects depend in part on the range of difficulty examined.

- The significant results for false alarm type across the three tasks suggest that these dimensions are not perceptually independent. Efforts are underway to investigate this possibility.

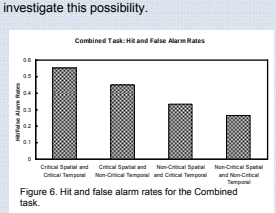


Figure 6. Hit and false alarm rates for the Combined task.

References

Available upon request

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, P.A. Hancock, *Principal Investigator*. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

Impact of Extraversion on Performance and Workload in Target Detection

J. E. Thropp, J. L. Szalma, J. M. Ross, and P. A. Hancock

University of Central Florida
Orlando, FL

The influence of operator extraversion on performance and workload in a target detection task was investigated. Extraversion predicted response bias in a task requiring temporal discrimination, but only in the presence of intermittent noise. These effects were not observed in tasks involving either spatial uncertainty or a combination of these spatial-temporal characteristics. Higher extraversion predicted reduced global workload in all three tasks, a relationship largely driven by lower reported levels of temporal demand and frustration. One possible explanation for these effects may lie in perceptions of time. Extraverts tend to focus attention to external cues, which can lead to the perception of time passing more rapidly. Introverts are more internally focused, which may lead to the perception of time passing more slowly. Hence, the internal clock of extraverted individuals may progress more rapidly, and more closely approximate the similarly rapid task pace. Conversely, introverts experience temporal distortion in the opposing orientation, which induces temporal demand due to the inverse relationship between a decelerated internal clock and a rapidly paced task.

Alternatives to Grouping Individuals in the Analysis of Individual Differences Data: Applications to Optimism and Pessimism

James. L. Szalma and Peter A. Hancock

There is growing recognition by applied experimental and human factors psychologists of the importance of considering operator characteristics in the design and operation of modern systems (e.g., Karwowski & Cuevas, 2003). It is therefore crucial that the techniques used to evaluate individual differences be sufficiently sensitive to detect subtle effects often found in this type of research. Most experimental psychologists and human factors professionals rely on ANOVA techniques to analyze individual differences data, which are often measured using continuous scales. The manipulated independent variables (e.g., training, display format) are usually categorical. Thus, the typical procedure for the analysis of the interaction of categorical and continuous variables is to categorize the continuous variable via median splits or extreme groups (e.g., quartile splits). While such an approach permits use of the ANOVA model, categorization of continuous variables results in a loss of information and a possible decrease in the sensitivity of tests of hypotheses regarding the effect of these variables on the dependent measures of interest (e.g., the performance, workload, and stress experienced by operators). In addition, categorizing individual differences data effectively removes the individual from individual differences research by forcing individuals who vary along a specified range of a continuous variable to be considered qualitatively equivalent. Individuals with meaningfully different scores (e.g. moderate extraversion versus high extraversion) may therefore be classified together as ‘high’ or ‘low’ on a trait. A further disadvantage is that as a result of an arbitrary cut score, a

difference of one unit can mean the difference between classification as ‘high’ or ‘low’ on a trait (i.e., median splits) or between inclusion or exclusion in an analysis (i.e., quartile splits used in extreme groups approaches).

In addition to influencing statistical results, these methods also influence the manner in which individual differences variables are interpreted and the precision with which these theoretical constructs are considered. For instance, because the ANOVA based techniques force individual differences into categories, interpretation and discussion of these variables tends to be phrased in terms of ‘group’ differences (e.g., extraverts versus introverts; pessimists versus optimists). This may have the effect of creating theoretical ‘categories’ that do not actually exist. In contrast, preserving the continuity of the data affords interpretation to better reflect actual characteristics of individuals by directing attention to the degree to which an individual is high or low on a trait rather than that individual’s assignment to a category.

Experimental psychologists and human factors researchers should not be constrained by these methods, however. Established regression techniques are available to analyze categorical and continuous variables, and their interaction, without arbitrary division of the continuous variable. This procedure involves dummy or effect coding the categorical independent variables and computing product vectors for interactions among those variables and the continuous (individual difference) variable. Hierarchical regression is used to test for main effects and interactions, with the latter determined by whether the addition of the product vector terms to the regression model significantly increments the variance accounted for after the main effects have been entered (see Pedhazur, 1997). Significant interactions obtained by this method can be further analyzed using the

Johnson-Neyman procedure for analyzing regions of significance (for a thorough list of references on this procedure, see Pedhazur, 1997). These analyses, analogous to tests of simple main effects in ANOVA, reveal the ranges along the individual differences variable in which significant group differences occur.

The regression-based approach described above allows for analysis of interactions between categorical and continuous variables without resorting to loss of information provided by the continuous variable. Note that the techniques described here are not new. For instance, the Johnson-Neyman procedure was originally derived in 1936, and in domains such as education this approach has been referred to as analysis of 'aptitude-treatment interactions' (see Cronbach & Snow, 1977). Despite this long history, application of these techniques to individual differences research has rarely extended into the realm of experimental psychology or human factors.

To illustrate these techniques and their application to individual differences research, the current paper reviews data from several experiments investigating the role of dispositional pessimism and optimism. Note that these studies employed a measure of optimism and pessimism that assesses these traits as partially independent dimensions, so that an individual can, for instance, be high in both optimism and pessimism (see Dember, Martin, Hummer, Howe, & Melton, 1989).

Using the regression techniques described above, these studies collectively indicate that these traits influence how observers respond to target detection tasks. Specifically, individuals higher in pessimism tend to experience higher levels of distress and less energy and motivation after a target detection task than prior to its start. In addition, those higher in pessimism tend to engage in more emotion focused and avoidant coping

strategies in response the stress imposed by these tasks, a pattern of results that indicates that high pessimism is maladaptive for coping with task-based stress (see Matthews & Campbell, 1998). However, these effects are to some degree task dependent.

References

Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington

Dember, W. N., Martin, S. H., Hummer, M. K., Howe, S. R., & Melton, R. S. (1989). The measurement of optimism and pessimism. *Current psychology: Research and reviews,* **8**, 102-119.

Karwowski, W. (Chair), & Cuevas, H.M. (Co-chair) (2003, October). Considering the importance of individual differences in human factors research: No longer simply confounding noise. Symposium panel discussion conducted at the 47th annual meeting of the Human Factors and Ergonomics Society, Denver, CO.

Matthews, G. & Campbell, S. E. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society, 42, 821-825.*

Pedhazur, E.J. (1997). *Multiple regression in behavioral research*, 3rd Ed. Ft. Worth: Harcourt Brace.

Alternatives to Grouping Individuals in the Analysis of Individual Differences Data: Applications to Dispositional Pessimism

J.L. Szalma and P.A. Hancock

University of Central Florida



Abstract

The typical procedure for the analysis of the interaction of categorical and continuous variables is to categorize the continuous variable via median splits or extreme groups (e.g., quartile splits). While such an approach permits use of the ANOVA model, and can be very fruitful for individual differences research, we argue that alternative approaches should be considered when one is interested in exploring the influence of a continuous individual differences variable on performance or operator state. Specifically, regression techniques are available to analyze categorical and continuous variables, and their interaction, without arbitrary division of the continuous variable. Although in many cases the same outcome is observed with both approaches, there are cases in which non-significant differences using a median split mask actual differences that are detected when a regression approach is used. An application of the regression technique to data on the relation between Pessimism and the stress and workload of sustained attention is presented.

Introduction

There is growing recognition by applied experimental and human factors psychologists of the importance of considering operator characteristics in the design and operation of modern systems (e.g., Karwowski & Cuevas, 2003). It is therefore crucial that the techniques used to evaluate individual differences be sufficiently sensitive to detect subtle effects often found in this type of research. Most experimental psychologists and human factors professionals rely on ANOVA techniques to analyze individual differences data, which are often measured using continuous scales. The manipulated independent variables (e.g., training, display format) are usually categorical. Thus, the typical procedure for the analysis of the interaction of categorical and continuous variables is to categorize the continuous variable via median splits or extreme groups (e.g., quartile splits). While such an approach permits use of the ANOVA model, categorization of continuous variables results in a loss of information and a possible decrease in the sensitivity of tests of hypotheses regarding the effect of these variables on the dependent measures of interest (e.g., the performance, workload, and stress experienced by operators). These limitations of categorization are demonstrated in Figure 1.

In addition to influencing statistical results, these methods also influence the manner in which individual differences variables are interpreted and the precision with which these theoretical constructs are considered. For instance, because the ANOVA based techniques force individual differences into categories, interpretation and discussion of these variables tends to be phrased in terms of 'group' differences (e.g., extraverts versus introverts; pessimists versus optimists). This may have the effect of creating theoretical 'categories' that do not actually exist. In contrast, preserving the continuity of the data affords interpretation to better reflect actual characteristics of individuals by directing attention to the degree to which an individual is high or low on a trait rather than that individual's assignment to a category.

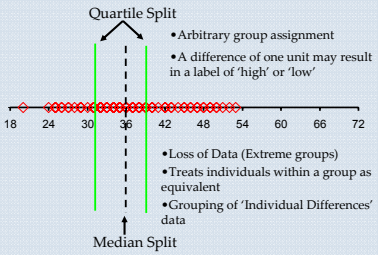


Figure 1. Effect of categorizing a continuous variable. The points are scores on the pessimism scale of the OPI. See text for details.

Alternative Approach: Regression

Established regression techniques are available to analyze categorical and continuous variables, and their interaction, without arbitrary division of the continuous variable. In this approach:

1) The categorical independent variable is coded using dummy or effect coding and product vectors are computed for interactions with the continuous individual differences variable. An example of dummy coding is shown in Table 1.

2) Hierarchical regression is used to test for main effects and interactions (Attribute-Treatment interactions; see Pedhazur, 1997).

Significant interactions obtained by this method can be further analyzed in one of two ways:

1) Using the Johnson-Neyman procedure for analyzing regions of significance (see Figure 2; for a thorough list of references on this procedure, see Pedhazur, 1997). This analysis, analogous to tests of simple main effects in ANOVA, reveal the ranges along the individual differences variable in which significant group differences occur.
2) Computing separate regressions on the individual difference variable for each group.

Note that the techniques described here are not new. For instance, the Johnson-Neyman procedure was originally derived in 1936, and in domains such as education this approach has been referred to as analysis of 'aptitude-treatment interactions' (see Cronbach & Snow, 1977). Despite this long history, application of these techniques to individual differences research has rarely extended into the realm of experimental psychology or human factors.

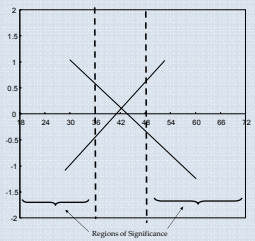


Figure 2. An example of regions of significance, denoted by the brackets. In this example, the scores for the two conditions (represented by different regression lines) would be significantly different only for scores lower than 36 or higher than 48.

To illustrate these techniques and their application to individual differences research, the current paper reviews data from several experiments investigating the role of dispositional pessimism. Each study measured pessimism using the Optimism-Pessimism Inventory (OPI; see Dember, Martin, Hummer, Howe, & Melton, 1989).

Example 1: The Effect of KR and Pessimism on Task Engagement in Vigilance

- Three forms of KR: Hit KR, False Alarm KR, and Miss KR
- Comparison to a No-KR control and a composite-KR control

Table 1. Dummy coding of categories of knowledge of results (KR) in a vigilance experiment.

	D_H	D_{FA}	D_M
Hit KR	1	0	0
FA KR	0	1	0
Miss KR	0	0	1
Comp KR	0	0	0

Regression Model:

Criterion/Dependent variable: Task Engagement (TE)

Predictors/Independent Variable: Pessimism (P), KR (D_H, D_{FA}, D_M), and Pessimism x KR ($D_{HxP} + D_{FAxP} + D_{MxP}$)

Hierarchical Regression

Model 1: $TE = P$

Model 2: $TE = P + (D_H + D_{FA} + D_M)$

Model 3: $TE = P + (D_H + D_{FA} + D_M) + (D_{HxP} + D_{FAxP} + D_{MxP})$

Table 2. Summary of Hierarchical Regression Analysis (N=110)

Model	ΔR^2	$F_{\Delta R^2}$
1	.03	3.35
2	.03	.95
3	.12	4.74*

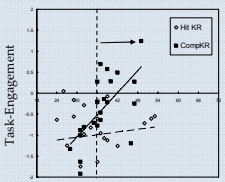
* $p<.05$

Table 3. Summary of tests of the regression co-efficients

Variable	b	t
P	.09	3.63**
D_H	2.57	2.13*
D_{FA}	5.25	3.80**
D_M	1.99	1.59
D_{HxP}	-.07	-2.27*
D_{FAxP}	-.14	-3.71**
D_{MxP}	-.06	-1.80

Note. The terms corresponding to the significant pessimism by KR interaction are highlighted in yellow. * $p<.05$

Figure 3. Pre-Post Vigil Task Engagement as a Function of Pessimism for Two KR conditions



Intersection = 29.87
Lower Bound: 36.67
Upper Bound: 313.61

Note. The vertical dotted line represents the threshold for the region of significance, as computed using the Johnson-Neyman procedure. The arrow points toward the region of significance.

Example 2: The Effect of Task type and Pessimism on Pre-Post Task Distress in Vigilance

Two tasks using configural displays: Focused attention (dot-distance) vs. integration (midpoint averaging task

Task type was dummy coded (0,1), and hierarchical regression was employed. Criterion variable: Distress (D); Predictors: Pessimism (P) and Task type (D_T).

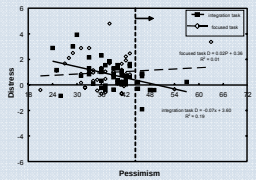
Regression Models:

Model 1: $D = P$

Model 2: $D = P + D_T$

Model 3: $TE = P + DT + D_{TxP}$

Figure 4. Pre-Post vigil Distress as a function of Pessimism for two Vigilance tasks (N=96)



Example 3: The Effect of Display type and Pessimism on Global Workload in Vigilance

Comparison of configural displays: well-mapped vs. poorly-mapped emergent features

Task type was dummy coded (0,1), and hierarchical regression was employed. Criterion variable: Distress (D); Predictors: Pessimism (P) and Task type (D_T).

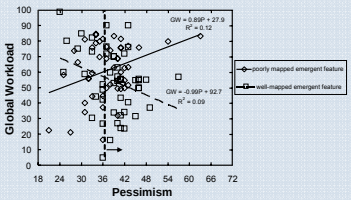
Figure 5. Global workload as a function of Pessimism for Well-mapped and poorly-mapped configural displays in Vigilance (N=32)

Regression Models:

Model 1: $D = P$

Model 2: $D = P + D_T$

Model 3: $TE = P + D_T + D_{TxP}$



More Complex Designs

- Multiple independent variables can be dummy coded (factorial designs)
- Effect coding can also be used
- More than one individual differences variable can be entered into the model
  - problem: complex interactions (e.g., 4-way interactions)

References

Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington

Dember, W. N., Martin, S. H., Hummer, M. K., Howe, S. R., & Melton, R. S. (1989). The measurement of optimism and pessimism. *Current psychology: Research and reviews, 8*, 102-119.

Karwowski, W., & Cuevas, H. M. (2003). Occupational ergonomics: Human diversity and individual differences in human factors research. *No longer simply confounding noise.* Symposium panel discussion conducted at the 47th annual meeting of the Human Factors and Ergonomics Society, Denver, CO.

Matthews, G. & Campbell, S. E. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society, 42*, 821-825.

Pedhazur, E.J. (1997). *Multiple regression in behavioral research.* 3rd Ed. Ft. Worth: Harcourt Brace.

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, Dr. Elmar Schmeisser, Technical Monitor (Grant# DAAD19-01-1-0621). The research was facilitated by a DARPA-funded program under Grant NBCH1030012, CMDR Dylan Schmorrow, Technical Monitor. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government.

Attitudes of Police Officers toward Seat Belt Use While on Duty

Stafford, S.C., Oron-Gilad, T., Szalma, J.L., Delasontos, K., & Hancock, P.A.

Seat belt usage has been estimated to be 40–50% effective in preventing fatalities (Partyka, 1988). Police officers on patrol duty spend a considerable portion of their time in their cruisers, and they are expected to comply with seatbelt laws and serve as an example for other citizens. However, it is possible that the typical seatbelt design interferes with police officers' operational work by lengthening response time to threatening situations. Thus, in high threat environments there may be a direct conflict between operational safety (effective responses to threat) and driver safety (seatbelt use). In this delicate balance between operational and driving safety there are many situations where the usage of the seat belt not only does not enhance the police driver's safety but also creates a higher risk level by not enabling quick maneuvers and easy access to the weapon.

There has been very little research regarding police officer driving habits, particularly in seat belt use. In regard to driving ability, Dorn and Brown (2003) have found that police officers do not perceive themselves as better drivers than others - on the contrary they claim that they need more specific driving training (e.g., night driving training). Thus, it may be that officers are not confident in seatbelt use when on duty, particularly in high threat environments. An initial examination of seat belt use by police officers, and their attitudes toward seatbelt use, was the primary goal for the present study.

Participants were 341 police officers who were members of police departments in municipalities in the southeastern U.S. Officers volunteered to participate in the study without compensation from the researchers. For the purposes of the present study, only those individuals classified as patrol officers were included for analysis (249 males and 41 females).

The questionnaire consisted of 61 items divided into three sections: 1) demographic information; 2) personal habits regarding seat belt use at home and work; and 3) opinions regarding seat belt use during a work shift. Questionnaires were administered to officers at the beginning of their shift by their shift commander. The questionnaire included a cover letter describing the general purpose of the study and assuring the participants that their responses would remain anonymous. Officers were instructed to complete the questionnaire by the end of their shift, and to deposit it in specified location. Officers did NOT return the questionnaire to their shift commander.

For the current study, only a portion of the results were analyzed to determine the effect of environmental threat and shift on officers' confidence regarding their ability to effectively perform their operational duties while wearing their seat belt. Threat was defined as high versus low crime areas, and three shifts were examined: day (starting from 5-9AM and ending between 3-6PM), mid-shift (starting from 12-2PM and ending between 8-11PM), and night shift (starting from 6-10PM and ending between 6AM and 9AM the next morning). A 2 (threat) by 3 (shift) mixed ANOVA with repeated measures on the first factor revealed a significant effect for threat, $F(1,287)=55.77$, $p<.001$, $f=.43$,

with officers reporting greater confidence in their ability to use their seatbelts in low threat (M=3.53) than in high threat (M=3.14) areas. A significant effect was also observed for shift, $F(2,287)=6.53$, $p<.05$, f=.25. Post-hoc tests using the Tukey HSD procedure indicated that officers reported greater confidence in seat belt use during a day shift (M=3.71) than in either the mid (M=3.10) or night (M=3.23) shifts. The means for the latter two shifts did not differ significantly from one another. The interaction between these factors was not statistically significant ($p>.05$).

These results support the hypothesis that a conflict between operational and driving safety exists. Further, these effects are exacerbated in high threat areas, and during later work shifts. The latter results are likely due to the fact that officers typically experience higher workload (e.g., more threatening dispatch calls, more events requiring emergency response) during the mid and night shifts than during the day shift. These data indicate that current seatbelt designs in police cruisers paradoxically represent a safety problem to police officers. An obvious solution is the design of seatbelts more suitable for police work. More suitable seatbelts may also be useful for other operational situations such as military vehicles (Naylor, 2003) or other work domains that include high threat environments.

References

Dorn, L., & Brown, B. (2003). Making sense of invulnerability at work – a qualitative study of police drivers. *Safety Science*, 41, 837-859.

Naylor, S.D. (2003). 'Heroes everywhere I went', *Army Times*, 63(44), 8.

Partyka, S.C., 1988. Lives saved by seat belts from 1983 through 1987. DOT HS 807 324. *National Highway Traffic Safety Administration*, Washington, DC.

What do we know about Teams? Toward A Practical Framework

Heather A. Priest, C. Shawn Burke, Eduardo Salas, & Clint A. Bowers

Psychology Department, University of Central Florida, Institute for Simulation & Training

Introduction

In order to remain adaptive and prosper in the emerging global and technological environment, many organizations are witnessing a flattening of traditional, hierarchical structures in favor of teams (Kozlowski & Bell, 2002; Zaccaro, Rittman, Orvis, Marks & Mathieu, 2002). Practitioners looking for guidance often find the team literature fragmented and narrow in its conceptualization of teams. For example, the team literature has often simplified the picture of teams by treating interdependence as a dichotomous variable. Additionally, given operational examples of 'real world' teams, we expect that not only do teams differ in interdependence, but that within team type there are sub-categories that differ as well (i.e., distribution characteristics). Existing task and/or team typologies tend to lack integration – traditionally focusing on either task or team characteristics, but rarely integrating the two. This makes it difficult to translate research findings into the practical guidance needed for the wide variety of teams that are appearing within organizations.

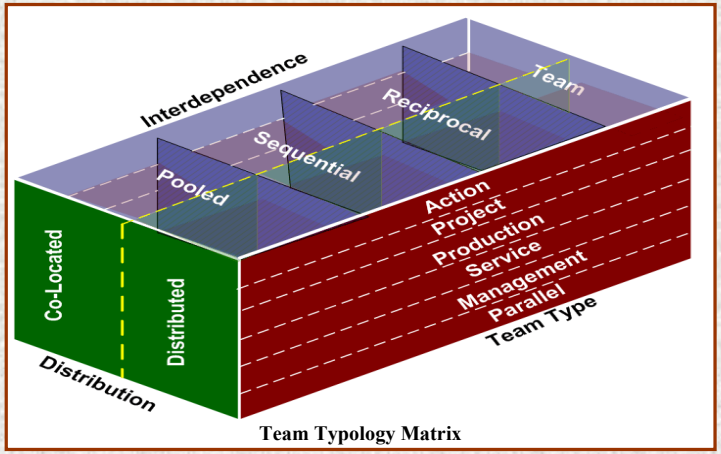
Goals

It is our hope that this framework will serve as food for thought and accomplish three additional goals.

1. To serve as a practical mechanism by which to systematically categorize what we really know about teams after forty years.
2. To facilitate the development of prescriptive guidance for those within organizations.
3. To facilitate the identification of future research needs in a systematic manner.

Proposed Approach

We propose that one way to increase the functionality of team typologies would be to devise a practical, integrative framework that includes team type, distribution characteristics, and interdependency as a continuous variable. The method for developing this framework was to use what we have gained to date regarding teams and interaction. The goal was to integrate several typologies on teams and interdependencies) to develop this tool. The framework consists of Sundstrom et al.'s team type (see Table 1) on one axis and Saavedra et al.'s levels of interdependency (see Table 2) on the other axis. An additional variable was included in this framework on the third axis: distribution. Little is known about what distribution does to traditional team interaction. Therefore, it is important for future research to further dissect the differences between distributed and co-located teams, across team type and interdependency level.



Team Typology Matrix

Existing Team Typologies

Table 1. Team Types (Sundstrom et al., 2000)

Team Type	Definitions
Action	Groups that conduct complex, time-limited performance events that involve audiences, challenging environments, or adversaries
Project	Teams that carry out projects that are time-limited, specialized, and defined; typically teams are cross-functioning and disband after finishing a project
Production	Teams who produce tangible products repeatedly and are on the frontline
Service	Teams that conduct transactions repeatedly with customers
Management	Teams that coordinate work of subordinates through budgeting, planning, policy making
Parallel/Advisory	Teams function across time, place, and structure and must create a shared core

Table 2. Team Interdependency (Saavedra et al., 1993)

Team Interdependency	Definitions
Pooled	Each member makes a contribution to the output without the need for direct interaction
Sequential	Team members have different roles and perform different tasks in a prescribed order. One team member must act before the other can
Reciprocal	Team members have different roles and expertise, use temporally logged 2 way interactions
Team	Team members jointly diagnose, problem solve, and collaborate, it is up to group discretion to decide the particular course that they take to achieve outputs/goals

Framework

This framework is manifested as an organizing matrix that is used as a heuristic to help classify team research based on multiple factors. The framework proposes a possibility of 48 teams. The variables on each of the 3 axis intersect at various points within the matrix. This framework will enable researchers to classify teams based on three separate categories, more accurately representing the complexity of teams in the real world. This has not consistently been done in the current team research environment. Most research on teams has been done on action teams with high levels of interdependence that are co-located. This tool can be used to not only guide future research, but as a mechanism to organize what is know and thereby identify what is not known. A future application, already being developed, is an effort to classify the team research from the 1980 through the present, based on multiple factors. Researchers can then evaluate the findings of the last 25 years and more systematically examine future research needs.

Questions: hpriest@ist.ucf.edu; References available upon request

What do we know about Teams? A Practical Framework

Priest, H. A., Burke, C. S., Salas, E. & Bowers, C. A.

The past forty years has witnessed much change in the areas of organizational theory, structure, and business practice. Specifically, technological advances, geopolitical stability, and free trade agreements have increased organizational competition within a global economy. In order to remain adaptive and prosper under these circumstances many organizations are witnessing a flattening of traditional, hierarchical structures in favor of teams (Kozlowski & Bell, 2002; Zaccaro, Rittman, Orvis, Marks & Mathieu, 2002). In fact, 80% of surveyed workers report they are currently members of at least one team and this estimate will continue to increase in step with evolving environmental complexities (Fiore, Salas & Cannon-Bowers, 2001).

With the increased use of organizational teams has come increased research. This research has taught practitioners and researchers alike much about teams. For example, (a) teams are characterized by interdependency, adaptive action, and common goals (Salas, Dickinson, Converse, & Tannenbaum, 1992), (b) both teamwork (i.e., behavioral interactions and attitudinal responses needed to coordinate) and taskwork (i.e., operational) skills are necessary (Morgan, Glickman, Woodard, Blaiwes, & Salas, 1986; Hackman & Morris, 1975; Bass, 1990; Brannick et al., 1993), and (c) teamwork is characterized by a complex set of knowledges, skills, and attitudes (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Salas & Cannon-Bowers, 2000).

Despite all that has been learned we have only begun to scratch the surface and much of the literature is disjointed and neglects to offer practical prescriptive guidance for practitioners. As a field we have a tendency to investigate the global aspects of teams, yet often fail to probe deeply into these complex entities. Therefore, we argue that it is time to return to basics and

thoroughly investigate some of the defining characteristics of teams. Combining the above two points brings us to the issue of team typologies, a potentially very useful practical tool, and interdependence.

Team typologies arose out of an awareness that organizations housed a variety of teams that came in many different shapes and sizes. In an effort to begin to identify how various organizational teams may be different several researchers have begun to delineate different types of teams that are found in organizations (e.g., Sundstrom, McIntyre, Halfhill, & Richards, 2000; Devine, 2002). While these team typologies have the potential to be very useful the predominant number are mere classification systems that offer no prescriptive guidance in terms of their practical use. In addition, while it is apparent that organizational teams possess different levels of interdependency the team literature and the corresponding team taxonomies have traditionally treated interdependence as a dichotomous variable, failing to recognize its continuous nature. One only has to look to operational teams to realize that teams possess different levels of interdependency (e.g., surgery teams, crisis response teams, combat teams, cockpit crews, management teams). In addition, taking a multi-disciplinary perspective and looking outside the team literature one finds several taxonomies that serve to classify interdependence as a continuous variable (see Saavedra, Earley, & Van Dyne, 1993; Thompson, 1967; Van de Ven, Delbecq, & Koenig, 1976).

There is no question that teams are interdependent and that up until now the team literature has tended to simplify the picture by treating interdependence as a dichotomous variable. Moreover, given what is known about the nature of teams, their defining characteristics, and operational examples of 'real world' teams we expect that not only do teams differ in their level of interdependence, but that within team type there are sub-categories that

differ in terms of interdependence (e.g., action teams: operating room teams, combat teams). Therefore, one way to increase the functionality of team typologies while at the same time 'returning to basics' and further expanding our knowledge would be to devise a practical, integrative framework that includes team type, task characteristics, and interdependency as a continuous variable. It is our hope that this framework will serve as food for thought and accomplish three additional goals. First, it will serve as a practical mechanism by which to systematically categorize what we really know about teams after forty years. Our hypothesis is that most of the research has been conducted on a limited sample of teams with similar levels of interdependency. Second, it will facilitate the development of prescriptive guidance for those within organizations. Finally, it will facilitate the identification of future research needs in a systematic manner.

The method for developing this framework was to use what we have gained to date regarding teams and interaction. The goal was to integrate several typologies on teams and interdependencies (e.g., Sundstrom et al., 2000; Saavedra et al., 1993; Devine, 2002; Hackman, 1990; Argote & McGrath, 1993; Steiner, 1972; Thompson, 1967; Van de Ven, Delbecq, & Koenig, 1976; Wagman, 2001) to develop a tool that will allow researchers and practitioners to systematically classify results and identify research needs. This tool would be manifested as an organizing framework that is used as a heuristic to help classify team research based on multiple factors. The framework consists of Sundstrom et al.'s team type on one axis and Saavedra et al.'s levels of interdependency on the other axis. The team types consist of action, project, production, service, and management teams (for a definition of these team types see Table 1). The typology of interdependencies developed by Saavedra et al. consists of pooled, sequential, reciprocal, and team interdependency (for a definition of these team interdependencies, see Table 2).

The cells of this framework consist of a classification of team research findings at a deeper level than what currently exists. What is missing in the literature is exactly this. A greater, deeper, richer, and more perceptive specification of what we know about how teams with different levels of interdependency function. This heuristic allows the systematic identification of how team type and interdependency combine to impact process and performance. In turn, this framework serves as a foundation for the development of more efficacious, practical guidance regarding teams.

References

Argote, L. & McGrath, J. E. (1993). Group processes in organizations: Continuity and change. In C. Cooper and I. T Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 333-389). Chichester, England: Wiley.

Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors, 40(4),* 672-679.

Brannick, M., Roach, R., & Salas, E. (1993). Understanding team performance: A multimethod study. *Human Performance, 6*, 287-308.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo, E. Salas, & Associates (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco, CA: Jossey-Bass.

Devine, D. J. (2002). A review and integration of classification systems relevant to teams in organizations. *Group Dynamics: Theory, Research, and Practice, 6(4),* 291-310.

Hackman, J. R. (Ed.) (1990). *Groups that work (and those that don't): Creating conditions for effective teamwork* (pp. 315-342). New York: Prentice-Hall.

Hackman, J. R., and Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: a review and proposed integration. In: Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology*, vol. 8. New York: Academic Press.

Morgan, B. B., Jr., Glickman, A. S., Woodard, E. A., Blaiwes, A. S., & Salas, E. (1986). *Measurement of team behaviors in a Navy environment* (Technical Report Number 86-014). Orlando, FL: Naval Training Systems Center.

Saavedra. R., Earley, P. C., and Van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology, 78(1),* 61-72.

Salas, E., & Cannon-Bowers, J. A. (2000). Teams in organizations: Lessons from history. In M. M, Beyerlein (Ed.), *Work teams: Past, present and future* (pp. 323-331). Netherlands: Kluwer Academic Publishers.

Steiner, I. (1972). *Group process and productivity.* New York: Academic Press.

Sundstrom, E. McIntyre, M., Halfhill, T., and Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics, 4(1),* 41-67.

Thompson, J. D. (1967). *Organizations in action.* New York: McGraw-Hill.

Thornton, R. C. (1992). *The effects of automation and task difficulty on crew coordination, workload, and performance.* Unpublished doctoral dissertation, Old Dominion University, Norfolk, VA.

Van de Ven, A.. Delbecq A., & Koenig, R. Jr. (1976). "Determinants of coordination modes within organizations." *American Sociological Review 41*, 322-338.

Table 1. Team Types (Sundstrom et al., 2000)

Team Type	**Definitions**
Action	Groups that conduct complex, time-limited performance events that involve audiences, challenging environments, or adversaries
Project	Teams that carry out projects that are time-limited, specialized, and defined; typically teams are cross-functioning and disband after finishing a project
Production	Teams who produce tangible products repeatedly and are on the frontline
Service	Teams that conduct transactions repeatedly with customers
Management	Teams that coordinate work of subordinates through budgeting, planning, policy making

Table 2. Team Interdependency (Saavedra et al., 1993)

Team Interdependency	**Definitions**
Pooled	Each member makes a contribution to the output without the need for direct interaction
Sequential	Team members have different roles and perform different tasks in a prescribed order; One team member must act before the other can
Reciprocal	Team members have different roles and expertise; use temporally lagged 2 way interactions
Team	Team members jointly diagnose, problem solve, and collaborate; it is up to group discretion to decide the particular course that they take to achieve outputs/goals

Patterns of Team Communication: Are all Teams Created Equal?






Heather A. Priest, Joseph Guthrie, C. Shawn Burke, Clint A. Bowers, Eduardo Salas, & Laura Milham

Department of Psychology
University of Central Florida
Institute for Simulation & Training



ACKNOWLEDGEMENT

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

UCF



Overview

- Overview
- Background
- Purpose/Hypotheses
- Method
- Results
- Discussion/Future Research



Background

- Teamwork is important…
- Communication is key for effective teamwork (e.g.,Thornton, 1992).
- Research hasn't scratched the surface (Bowers et al., 1998).



Background

- Typically, team communication analysis has examined:
  - Teamwork behaviors
    - Frequency
    - Quality
  - This is not enough…must go deeper


NBC




HOTEL

ST. LOUIS
WOLVES





Background

Just communicating is not enough…

Communication must be clear and understood to be effective!



Background

- **Exploratory Sequential Data Analysis**
  - Bowers et al., 1998 Results showed a general improvement gleaning information not available through frequency counts
- **Generalizability of results across teams?**

Category	Defined as:
Plan	A statement that communicates an expected or anticipated plan, event, goal, etc.
Action/ Command	A statement/command that outlines an action (e.g., gives directions)
Acknowledgement	A one-bit statement that confirms the receipt of a command or fact
Response	A statement in response to a plan, action, uncertainty
Uncertainty	A statement of uncertainty expressing a question that seeks information from the other team member
Facts	A statement about the environment or other fact



Purpose

- To examine patterns of team communication.
- To determine if there are different patterns across team type.



Hypotheses

- Teams will differ in frequency of what communications they use (e.g., command) based on their "type"
- Teams will differ in their communication sequences
  - Distributed teams will need to use closed loop sequences more often than co-located teams
  - Teams with higher levels of interdependency will exhibit more closed loop communication



Method: Participants

- Participants
  - 154 undergraduate students (77 two-person teams)
  - Randomly assigned to teams and conditions



Method: Design

- 3 (team type) x 2 (workload) between subjects design
- Team Type: Participants performed as members of one of three unique team types
  - Command-and-control,
  - Reconnaissance
  - Combat



Method: Design

- **Workload: manipulated by creating two levels (low, high) via the use of an auditory dual task configuration**
  - **Low workload**
  - **High workload**
    - **“Charlie/Delta, What is your position?”**
    - **Charlie/Delta had to respond where they were in the “virtual” world based on maps and contextual cues (e.g., buildings, landmarks)**



Method: Measures

- Measures taken pre- and post-test:
  - Dundee Stress State Questionnaire
  - Motivated Strategies for Learning Questionnaire, MSLQ
- Participants were given the NASA-TLX following the performance mission



Method: Communication Analysis

- All communication between team members during task was recorded.
- 6 coders were trained to get acceptable inter-rater reliability
- Coded communication according to Bowers et al., 1998 procedure



Background

- Open
  - Start: action, plan, uncertainty
  - End: not with acknowledgement or response
- Closed
  - Start: action, plan, uncertainty
  - End: acknowledgement

Category	Defined as:
Plan	A statement that communicates an expected or anticipated plan, event, goal, etc.
Action/ Command	A statement/command that outlines an action (e.g., gives directions)
Acknowledgement	A one-bit statement that confirms the receipt of a command or fact
Response	A statement in response to a plan, action, uncertainty
Uncertainty	A statement of uncertainty expressing a question that seeks information from the other team member
Facts	A statement about the environment or other fact

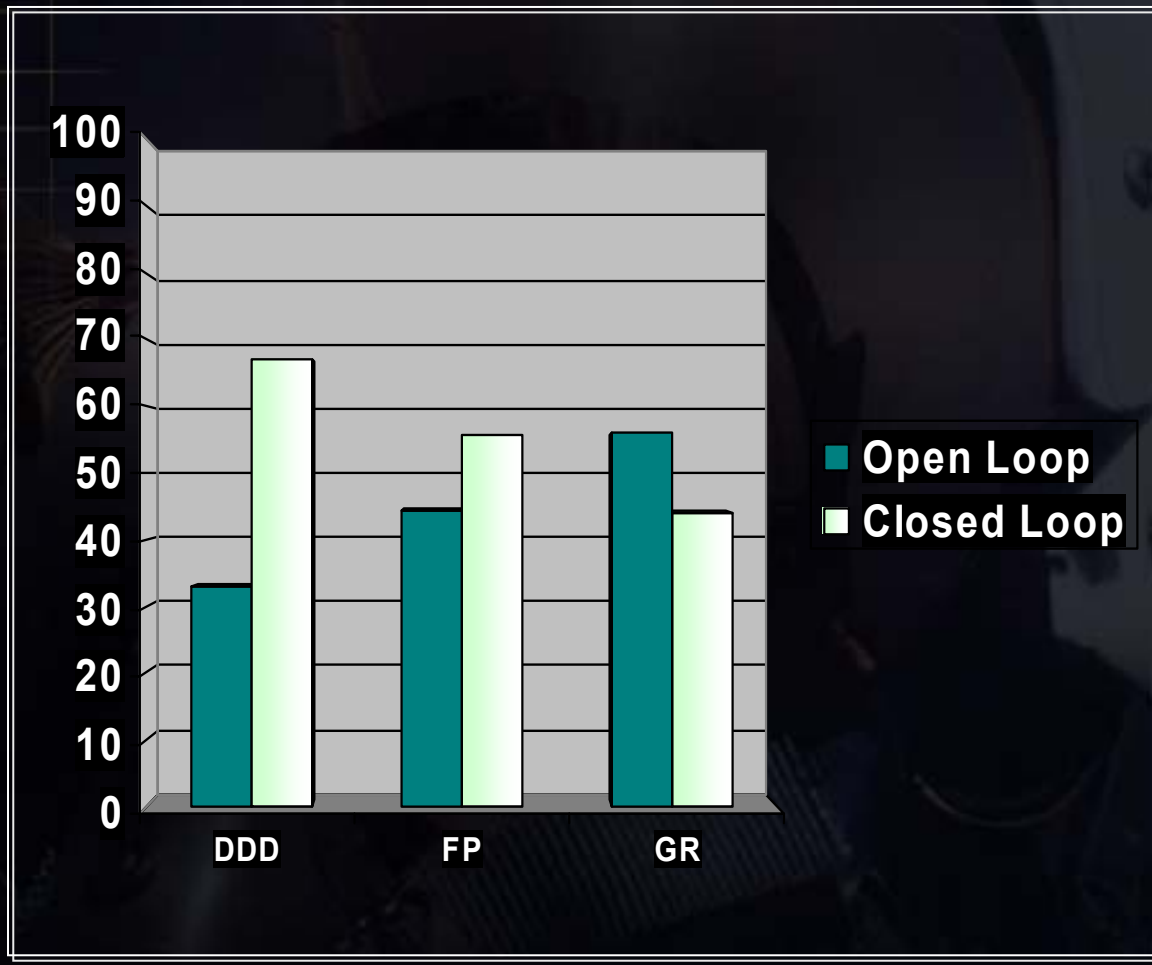


Results: Stress

- Teams in high workload focused on their primary task, offloading the secondary task.
- Thereby reducing their communication load.

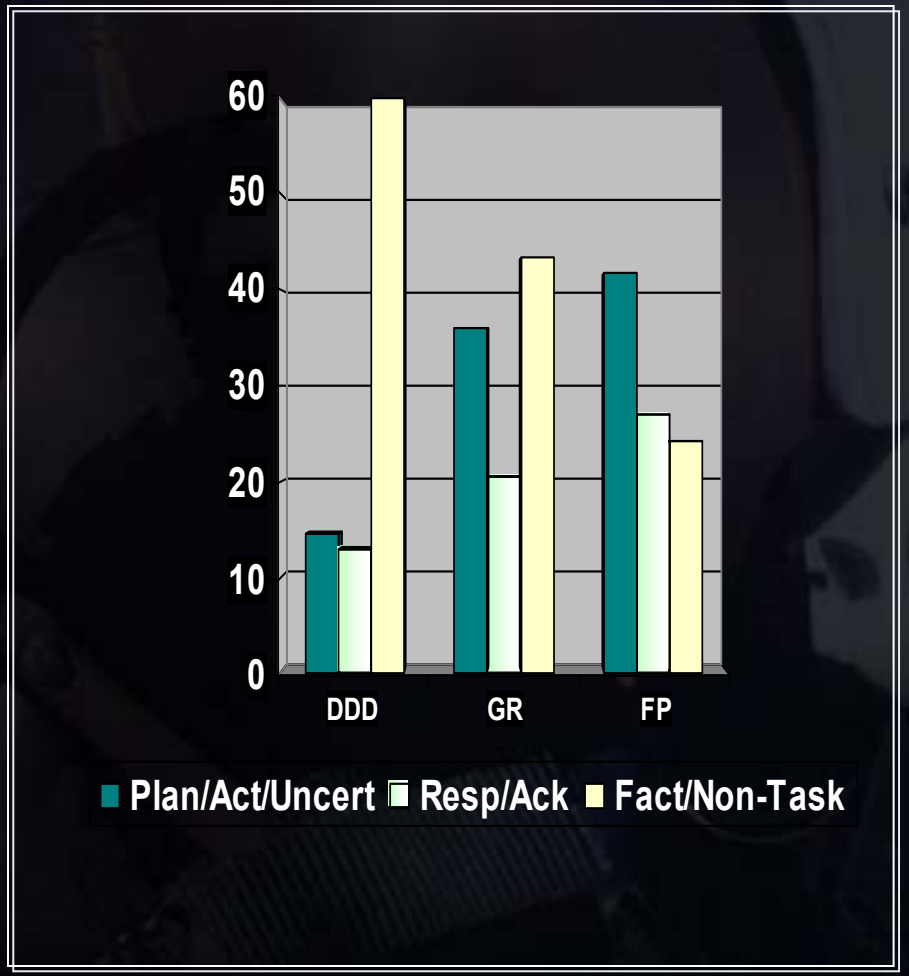
Results

- Found differences between teams in open versus closed loop communication sequences
  - DDD 66.9%
  - Flashpoint 55.7%
  - Ghost recon 44.1%



Results

- Significant differences found based on the type of communication used
  - DDD: facts or non-task related statements were used most to communicate
  - Ghost Recon: facts and non-task statements to communicate were used most, but open loop statements were also used often
  - Flashpoint: Planning/Action and uncertainties were used most, while responses and acknowledgements were also used often





Discussion

- Results support the hypothesis that different teams have different communication patterns
  - Indicating that teams have different communication needs in order to perform effectively
- Results also indicate that teams use different types of communication to accomplish their goals



Implications

- All teams may not be equal…
  - Training
  - Models of team effectiveness
  - Generalizability of findings



Implications

- Researchers and practitioners must …
  - More carefully define the teams they use
  - Probe deeper into issues regarding team type to examine what findings generalize and what are type specific
    - Is there a ‘team g’ or do we differentiate team type based on specific characteristics (e.g., interdependency)?
  - Must match theoretical predictions and empirical evidence with practice
    - Stop assuming that there is a ‘team g’ without investigation



Future Research

- Further examination of teams, team training, and how we define teams is needed
- Models of team effectiveness and training must be examined to determine their applicability across team type
- Need to evaluate non-verbal as well as verbal

Bottom Line…

- It is important to fully investigate what types of communication make teams successful
- Different types of teams require different types of communication to perform successfully
- It is easy to misunderstand both verbal and nonverbal cues…



The Pattern of Team Communication across Team Type: Are all Teams Created Equal?

Heather A. Priest, Joseph Guthrie, C. Shawn Burke, Clint Bowers, Eduardo Salas, & Laura Milham

The focus of traditional team research has revolved around performance and the measurable tangibles that teams deliver. However, within recent years there has been a move to understand the 'black box' known as team process. Typically this has been done through analyzing team communication in terms of frequency or quality and relating this to team performance. While insightful this method has yielded mixed results (Orasanu, 1990; Thornton, 1992). In an effort to further understand team process Bowers, Jentsch, Salas, and Braun (1998) applied elements of Exploratory Sequential Data Analysis (ESDA), a technique often used by researchers in other domains (Sanderson & Fisher, 1997) to the analysis of *communication sequences*. Results supported a general improvement over traditional communication analysis, gleaning information not available through frequency counts, but left a need to examine the generalizability of results across teams.

Although issues of generalizability exist across most domains within the teams literature it is especially troublesome as the majority of team research has tended to focus rather narrowly on a small subset of team types (e.g., command-and-control) with certain characteristics (e.g., co-located). Indicative of this trend is the fact that the most used testbeds in team research are DDD (Dynamic Distributed Decision-Making, Kleinman, Pattipati, Luh, & Serfaty, 1989), TIDE² (Team Interactive Decision Exercise for distributed expertise, Hollenbeck, Sego, Ilgen, Major, Hedlund, & Phillips, 1997), and TANDEM (Tactical Naval Decision Making System; Kozlowski, 1996), all of which are

Submitted to: the 112th Annual Convention of the American Psychological Association Honolulu, Hawaii, July 28 to August 1, 2004.

representative of command-and-control teams. These results are typically taken and applied to any collective classified as a team according to the broad definition accepted by today's team researchers: two or more individuals interacting interdependently and adaptively towards a common goal (Salas, Burke, & Cannon-Bowers, 2000) .

In an effort to begin to investigate how communication sequences might differentially relate to team performance across teams with different characteristics we created a testbed that would enable investigation of this question. Specifically, following the rationale depicted taxonomic literature on team type it was our belief that teams having different performance arrangements and interdependencies would see differential patterns of communication. In addition to examining differences in communication across teams we also sought to extend Bowers et al.'s (1998) work on ESDA.

Method

Participants were 154 undergraduate students (77 two-person teams) at a large southeastern university that were randomly assigned to teams and conditions. Researchers used a 3 (team type) x 2 (workload) between subjects design where participants performed as members of one of three unique team types (command-and-control, reconnaissance, combat) under one of two workload conditions. Based on the literature discussed earlier, researchres created 3 testbeds that would elicit 3 'team types' based on level of interdependency (Saavedra et al., 1993), members being co-located/distributed, and degree of shared awareness offered through the environment . Specifically, we used DDD to represent a command-and-control team (reciprocal interdependence, distributed expertise, moderate shared awareness) and then modified off-the-shelf computer games (COTS) to create a testbed which represented a combat

team (Ghost Recon; team interdependence, co-located, high shared awareness) and one which represented a search and rescue team (Flashpoint; sequential interdependence, distributed, low shared awareness). Missions, task, and team characteristics within each testbed were adapted so that they would be representative of the three above mentioned team types. The second between subjects variables, workload, was manipulated by creating two levels (low, high) via the use of an auditory dual task configuration.

After completing informed consent and background questionnaires, participants received task related training. At the completion of training participants completed an adaptation of the Dundee Stress State questionnaire (Matthews, Falconer, & Campbell, 1997) and the Motivated Strategies for Learning Questionnaire, MSLQ (Pintrich, Smith, Garcia, & McKeachie, 1993). Participants then completed their performance mission(s) and once again completed the Dundee and MSLQ. Participants completed the NASA-TLX and were debriefed as to the purpose of the study. Comparable performance measures were established across testbeds and communication during performance was recorded for analysis.

Results

There were slight differences between testbeds in open versus closed loop communication sequences (see figure 1). DDD produced the most closed loop communications, with 66.9% of their communication resulting in closed loops. Flashpoint produced a more even distribution but still had more closed loop communication, with 55.7% of their total communications resulting in closed loop. Ghost recon was the only testbed that produced more open loop communications, with only 44.1% of their total communications resulting in closed loops.

Submitted to: the 112th Annual Convention of the American Psychological Association
Honolulu, Hawaii, July 28 to August 1, 2004.

However, when examining communications of the teams on a deeper level, more significant differences are found based on the type of communication they used (see figure 2). DDD participants almost exclusively used facts or non-task related statements to communicate. In addition, ghost recon and flashpoint participants used a large percentage of facts and non-task statements to communicate, while flashpoint participants also used uncertainties and responses a majority of the time.

Conclusion

The results of the current study support the differences between team type. While there were no consistent patterns, the differences found in high performing and low performing teams across team types, support the notion that different teams have different communication needs in order to perform effectively. The implications of this research are widespread. For example, based on our results, it may not be appropriate to apply a general team finding (e.g., more communication is better) based on research using one team type (e.g., command and control) to another team type (e.g., a recon army team). The results reported here support further examination of teams, team training, and how we define teams. If there are not the overarching rules for effective teams that we have always relied on when it comes to communication, researchers and practitioners must more carefully define the teams they use and must move beyond the general team training typically applied regardless of team type. Future research must examine the both team models of effectiveness and team training to determine their applicability across team type. The findings based on communication analysis lead us to question: Is there a 'team g' or do we differentiate team type based on specific characteristics (e.g., interdependency)?

Submitted to: the 112th Annual Convention of the American Psychological Association
Honolulu, Hawaii, July 28 to August 1, 2004.

Acknowledgement

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

References

Hollenbeck, J.R., Sego, D.J., Ilgen, D.R., Major, D.A., Hedlund, J., and Phillips, J. (1997). Team decision making accuracy under difficult conditions: Construct validation of potential manipulations and measures using the TIDE² simulation. In M. Brannick, E. Salas and C. Prince (eds.), New Directions in Team Measurement, pp. 111-136. San Francisco: Jossey Bass.

Kleinman, D., Pattipati, K. Luh, P., and Serfaty, D. (1992). Mathematical models of team performance: A distributed decision-making approach. In R. Swezey and E. Salas, (Eds.)

Teams: Their Training and Performance (177-218). Norwood, NJ: Ablex Publishing Company.

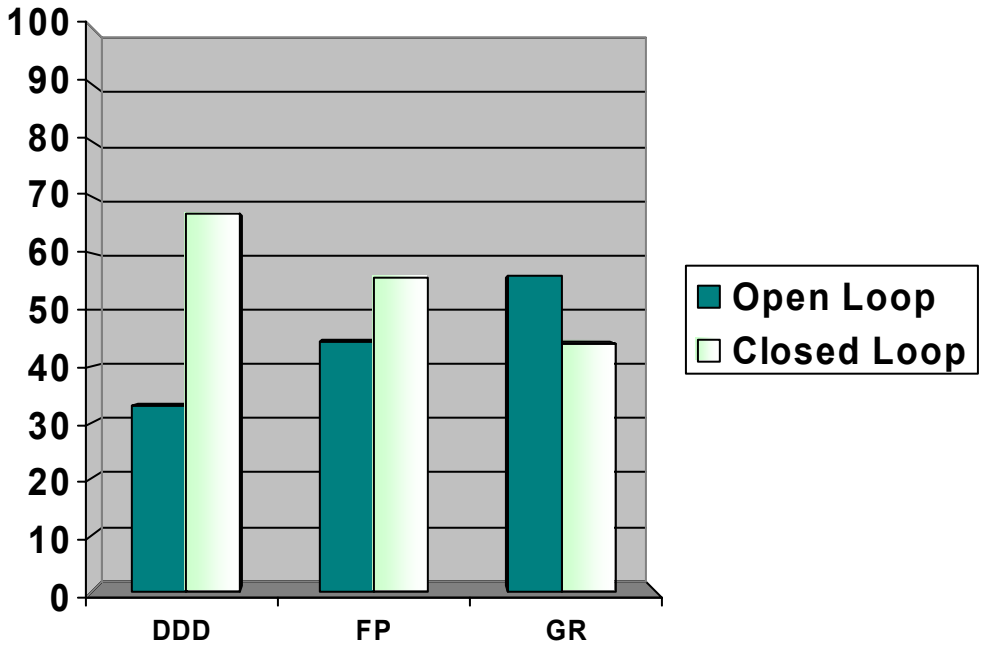
Kozlowski, SWJ (1996, March). TEAMS / TANDEM: An experimental platform for examining skill acquisition, adaptability, and effectiveness at individual and team levels of analysis. Presented at the Personnel Human Resource Research Group, University of Florida, Gainesville, FL.

Matthews, G., Falconer, S., & Campbell, S. (April 1997). *Assessment of Task Induced Stress.* Paper presented at BPS conference. Edinburgh, Scotland.

Submitted to: the 112th Annual Convention of the American Psychological Association Honolulu, Hawaii, July 28 to August 1, 2004.

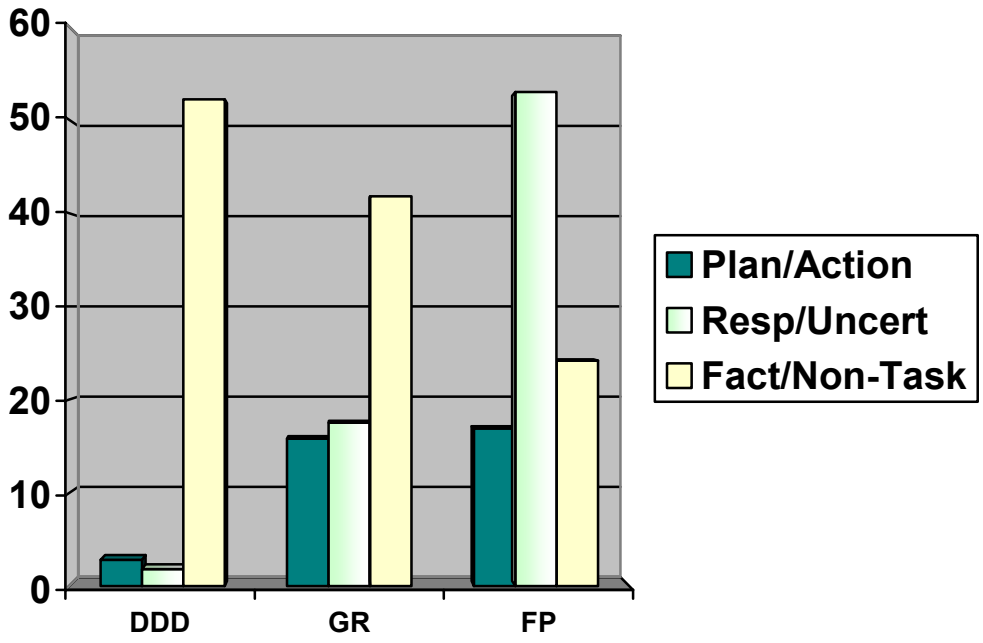
Pintrich, P.R., Smith, D.A.F., Garcia, T., & McKeachie, W.J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53, 801-813.

Figure 1. Closed loop versus Open Loop Communication by Testbed



Submitted to: the 112th Annual Convention of the American Psychological Association
Honolulu, Hawaii, July 28 to August 1, 2004.

Figure 2. Percentage of Overall Communication Used Based on Type



Submitted to: the 112th Annual Convention of the American Psychological Association
Honolulu, Hawaii, July 28 to August 1, 2004.

Development of COTS Testbeds:

Examining the Performance of Teams Under Stress

Guthrie, Jr. J.W., Priest, H., Asberg, K.,

Burke, C.S., Salas, E. & Bowers, C.A., University of Central Florida

Research on teams over the last decade has, in part, attempted to determine the effects of different stressors, such as time pressure, increased workload, ambiguity, noise and fatigue, on performance and how these effects are similar to and different from the effects on individuals. The majority of the research focused on teams and the factors affecting team performance has been conducted under the assumption that all teams are created equal, while other studies (Saavedra, Earley & Van Dyne, 1993) have focused on the areas in which teams may be different.

Specifically, Saavedra et al. (1993) described four different levels of team interdependency, which effects how team members interact. In teams with pooled interdependence each member makes a contribution to the output without the need for direct interaction. In teams with sequential interdependence, members have different roles and perform different tasks in a prescribed order. One team member must act before the other can work much like an assembly line. This requires more interaction, but the workflow is unidirectional. In teams with reciprocal interdependence members have different roles and expertise similar to sequential interdependence only the workflow is bidirectional. In teams with team interdependence members jointly diagnose, problem solve and collaborate; it is up to group discretion to decide the particular course that they take to achieve outputs/goals

In addition to teams having different levels of interdependency, teams also have different roles, tasks and available cues based on the type of team. Teams can either be distributed, where

team members are in different locations to complete the task, or co-located, where team members are together to complete the task. However, teams that are distributed or co-located can be further categorized based on their specific function. For example, command and control teams and reconnaissance teams are both distributed, but because of the different functions, different available cues and different roles have different levels of team interdependency.

Based on these differences in teams, research should then strive to reflect the variety of team roles, task roles, and team interdependencies found in real world settings. However, a recent search found that a majority of team studies primarily used one of three testbeds, Distributed Dynamic Decision-Making (DDD; Kleinman, Pattipati, Luh, and Serfaty, 1989), a simulation of distributed multi-person command and control teams in complex environments, Tactical Naval Decision Making System (TANDEM; Kozlowski, 1996), which uses command and control teams in simulating military decision-making environments, industrial environments, and health care applications, and Teams Incorporating Distributed Expertise ($TIDE^2$; Hollenbeck, Sego, Ilgen, Major, Hedlund, & Phillips, 1997), which uses command and control teams in order to help define 1) who does what tasks, 2) when they do them, 3) what workload the tasks impose, and 4) how the team coordinates its execution of the mission's tasks. Although all three testbeds used command and control teams, the results from such studies are, many times, applied to teams in general.

The Team Effectiveness Model (TEM; Tannenbaum, Beard, & Salas, 1992) provides a framework for team complexities that effect team process and performance. This framework takes into account different aspects of the task, complexity, structure and type, and the characteristics of the team, interdependence, cohesiveness and type, to determine the team processes and ultimately team performance. The assumption made by this framework is that

differences in the task and the team characteristics requires the team to use different processes in order to produce the desired outputs at a high level of performance. Teams using different processes would then require different training based on the specific processes used.

Based on the majority of previous research using the same types of teams and the likelihood that, based on TEM, different types of teams use different processes, our solution is to modify computer-off-the-shelf (COTS) games to provide different levels of interdependencies, task roles, and cues. These new testbeds would allow the comparison of new testbeds and existing testbeds (i.e., DDD) to test team differences based on interdependency, team role, and team task, to examine effects of stress (e.g., time pressure, workload) on different types of teams and to select tools for team measurements to be used across testbeds (e.g., NASATLX, sequential communication analysis).

For this goal to be realized, the identification of appropriate testbeds was necessary. The initial step in this process was to determine the requirements of a testbed. These characteristics generally fell under one of two categories: a) external validity, and b) practicality/internal validity. For a testbed to be externally valid, it needed to be able to tap into many of the required competencies for the target population, which in this case was the U.S. Army.

For example, the concept of situation awareness in team members must be readily available for measurement and observation with an appropriate testbed. In terms of testbed requirements, this translates into allowing for 1st person view or 1st level experience during testing. In addition, behaviors inherent in teams, such as team coordination and communication must be taken into account and allowed for by the testbed. The testbed structure should also be conducive to assessing other important concepts such as (team-) adaptability, navigation and

decision-making. Finally, to be externally valid, an appropriate testbed requires utilization of a shared mental model among team members.

To that end two COTS games were chosen for testbed development, Ghost Recon™ and Operation Flashpoint™. These games met the criteria for the type of testbed needed for this research. Both games were military based 1st person view action/strategy games. Potentially the most important aspect of the games is the game editor, which allows a user to create specific mission based scenarios within the structure of the game. In addition, these scenarios allowed us to create teams with different tasks, interdependencies, roles and cues. For Ghost Recon™, this allowed us to create a mission involving four objectives, using two-member co-located teams with team interdependence similar to a special forces combat team in the Army. In Operation Flashpoint™, we created two-member distributed teams with sequential interdependence similar to an Army reconnaissance team.

At the time of this paper, research has been conducted using the Ghost Recon™ testbed, the Operation Flashpoint™ testbed and DDD involving the impact of different types of stress, such as time pressure and workload, on team performance. This research primarily focuses on the potential process differences, investigated by looking at differences in communication, between the three different types of teams and the impact these differences might have on how we define teams and more importantly team training.

Acknowledgements

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research

Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

References

Hollenbeck, J.R., Sego, D.J., Ilgen, D.R., Major, D.A., Hedlund, J., and Phillips, J. (1997). Team decision making accuracy under difficult conditions: Construct validation of potential manipulations and measures using the TIDE^2 simulation. In M. Brannick, E. Salas and C. Prince (eds.), New Directions in Team Measurement, pp. 111-136. San Francisco: Jossey Bass.

Kleinman, D., Pattipati, K. Luh, P., and Serfaty, D. (1992). Mathematical models of team performance: A distributed decision-making approach. In R. Swezey and E. Salas, (Eds.) Teams: Their Training and Performance (177-218). Norwood, NJ: Ablex Publishing Company.

Kozlowski, SWJ (1996, March). TEAMS / TANDEM: An experimental platform for examining skill acquisition, adaptability, and effectiveness at individual and team levels of analysis. Presented at the Personnel Human Resource Research Group, University of Florida, Gainesville, FL.

Saavedra, R., Earley, P.C. & Van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology, 78(1)*, 61-72.

Tannenbaum, S.I., Beard, R.L. & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. In K. Kelley (Ed.), *Issues, theory and research in industrial/organizational psychology* (pp.117-153). Amsterdam: Elsevier.

Submitted for presentation at 2004 APA Annual Meeting, July 28-31, Honolulu, HI

SOLUTION: Modify computer-off-the-shelf (COTS) games to provide different levels of interdependencies, task roles, and cues.

•Use new testbeds (i.e., ghost recon, flashpoint) and existing tested (i.e., DDD) to test team differences based on interdependency, team role, and team task: Is there a team g?

•Examine effects of stress (e.g., time pressure, workload) on different types of teams.

•Selection of tools for team measurements to be used across testbeds (e.g., NASATLX, sequential communication analysis).

MURI-OPUS
Operator Performance Under Stress

University of Central Florida

Development of COTS Testbeds: Examining the Performance of Teams Under Stress



J.W. Guthrie, Jr., K Asberg, H.A. Priest, C. S. Burke, E. Salas, & C.A. Bowers

Department of Psychology

Institute for Simulation & Training University of Central Florida

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Overview

- Previous team research
- How teams are different
- Development of testbeds
- Applications

Previous Research

- Teams: Unique entities characterized by two or more individuals interacting socially and adaptively to achieve shared goals

- Teamwork occurs within and across multiple interdependent recursive episodes during action and transition stages of performance (Marks, Mathieu, & Zaccaro, 2001)

Previous Research

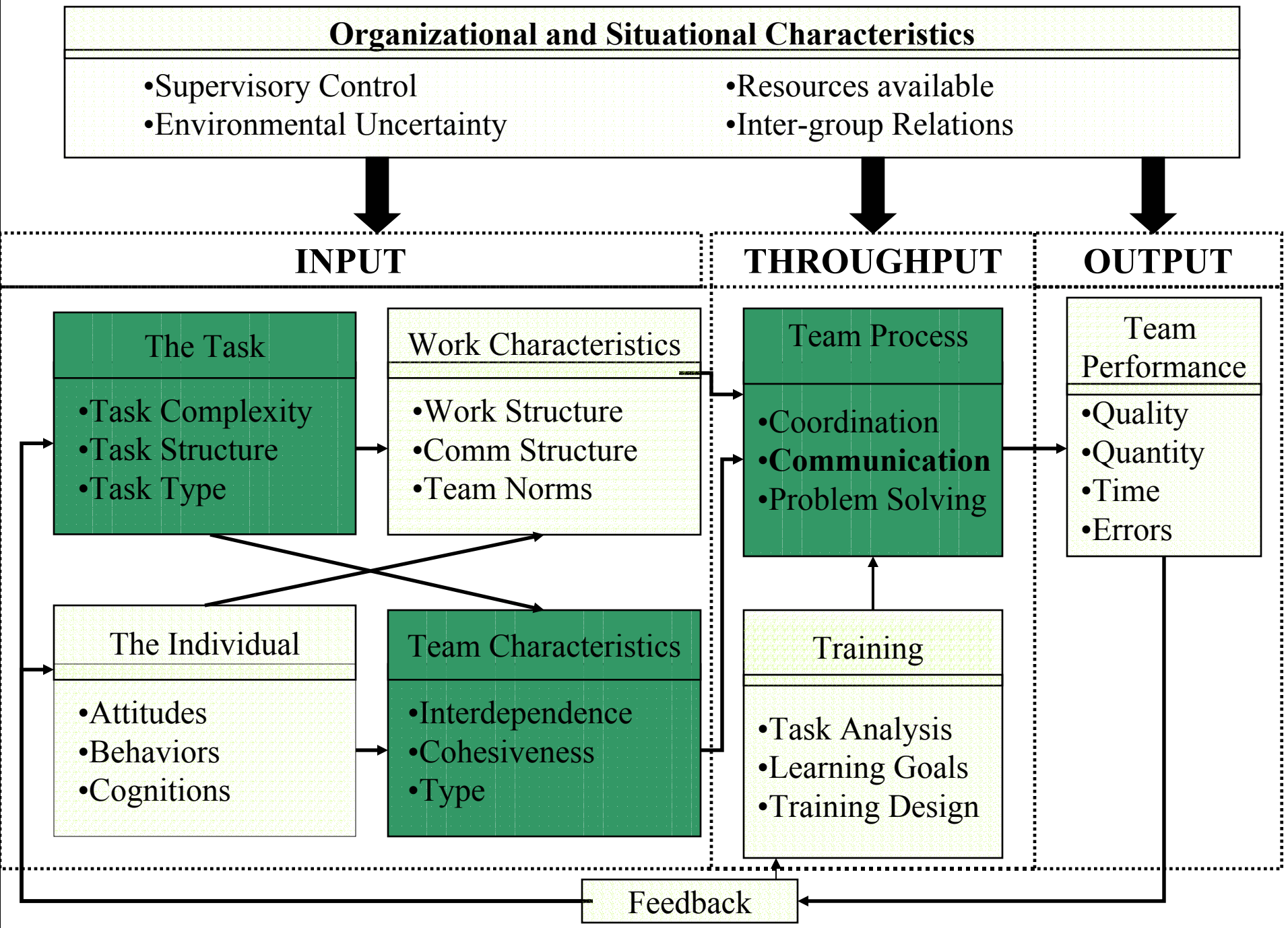
- A recent search found that a majority of team studies primarily used one of three testbeds:
  - TANDEM
  - TIDE² (Team Integrated Design Environment)
  - DDD (Distributed Dynamic Decision-Making)

Previous Research

- TANDEM
  - Uses command and control teams to simulate military decision-making environments (e.g., Joint Task Force, AWACS), industrial environments (e.g. manufacturing systems, civilian search and rescue) and even health care applications (e.g. distributed diagnosis).
- TIDE
  - Uses command and control teams to help define 1) who does what tasks, 2) when they do them, 3) what workload the tasks impose, and 4) how the team coordinates its execution of the mission's tasks.
- DDD
  - Distributed multi-person simulation and software tool for understanding how high-performance teams operate in complex environments

Previous Research

- Using primarily one type of team (command and control) throughout team research is a rather narrow concept of teams
- From the literature, we know that teams differ along many aspects…


Organizational and Situational Characteristics
•Supervisory Control
•Environmental Uncertainty
•Resources available
•Inter-group Relations
INPUT
THROUGHPUT
OUTPUT
The Task
•Task Complexity
•Task Structure
•Task Type
Work Characteristics
•Work Structure
•Comm Structure
•Team Norms
Team Process
•Coordination
•Communication
•Problem Solving
Team Performance
•Quality
•Quantity
•Time
•Errors
The Individual
•Attitudes
•Behaviors
•Cognitions
Team Characteristics
•Interdependence
•Cohesiveness
•Type
Training
•Task Analysis
•Learning Goals
•Training Design
Feedback

Testbed Development

- Our solution is to modify computer-off-the-shelf games (COTS) games to provide different levels of interdependency, task roles, and cues.

Testbed Development

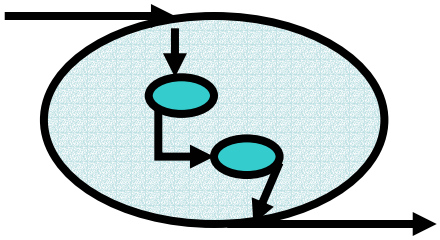
- Requirements of a testbed:
  - Needs to tap into many of the required competencies for the target population, which in this case was the US Army.
  - Ensure requirements for team-based testbed
  - Processes & performance indicators

Testbed Development

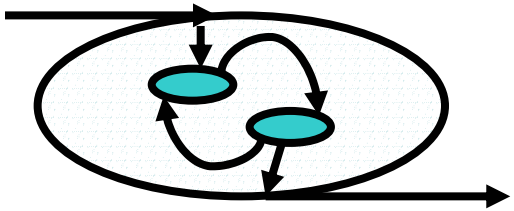
- Our search for appropriate testbeds resulted in the identification of two COTS that could be appropriately modified to fit our needs:
  - Operation Flashpoint
  - Ghost Recon

Testbeds

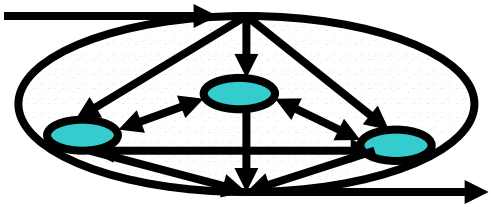
Sequential



Reciprocal

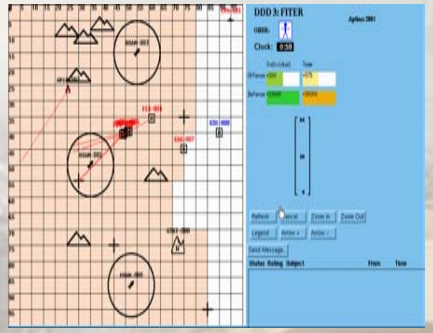


Team





Operation Flashpoint

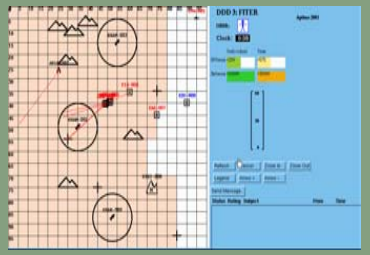


DDD



Ghost Recon

Team Types

Characteristics	Flashpoint™	DDD	Ghost Recon
			
Task Interdependence (Saavedra, Earley, & Van Dyne, 1993)	Sequential	Reciprocal	Team
Available cues of team mates' actions	Verbal Only	Verbal and Low Detail Visual	Detailed Visual and Verbal
Similar to:	Reconnaissance Team	Command and Control Team	Special Forces Combat Team

Future Research/Applications

- Allows for research that focuses on the potential process differences existing between different types of teams
- Can examine the effects of stress on different types of teams.
- Findings can be applied not only to how we define teams but more importantly how we train teams.

Differences in Teams

- Saavedra, Earley, & Van Dyne (1993)
  - Four levels of team interdependency
    - Pooled
    - Sequential
    - Reciprocal
    - Team

Differences in Teams

- In addition to having different levels of interdependency, teams also have different roles, tasks and available cues based on the type of team.
  - Co-located teams
  - Distributed teams
    - Differences within distributed teams

The Influence of Dispositional Optimism and Pessimism on Task Engagement for Spatial and Temporal Discrimination

James L. Szalma, Jennifer M. Ross and Peter A. Hancock

University of Central Florida

The Hancock and Warm (1989) model of stress and performance explicitly recognized that tasks themselves represent a major proximal source of stress for operators, and that task characteristics relevant to the stress-performance relation vary along two dimensions: 1) information structure, a spatial characteristic; and 2) information rate, a temporal characteristic. One possible mechanism by which stress impacts performance is via distortion of perceptions along these two dimensions, often manifested as attentional narrowing, and it may be that joint distortions of spatial and temporal stimuli occur as a result of declines in a common resource capacity for these dimensions (Hancock, Szalma, and Weaver, 2002).

Recent evidence suggests that the stress experienced by participants is influenced by the observer's dispositional pessimism and optimism (Thropp, Szalma, Ross, & Hancock, 2003). In addition, Thropp et al. reported that these effects depend on the spatial and temporal properties of the task. The current study was designed to further explore these effects by determining the effect of conjunctions of spatial and temporal discriminations and dispositional pessimism and optimism on detection performance and self reports of stress.

Participants in this experiment were twenty-three female and twenty-three male undergraduates who received financial compensation for their participation. Participants

in the noise group an 85 dBA intermittent white noise was presented via headphones. The ambient sound level in the quiet condition was 39.5 dBA.

All participants engaged in three tasks requiring discrimination of a line that varied along two of three dimensions: spatial, temporal, and luminance. The tasks differed according to spatial and temporal properties of the stimulus. The 'spatial-only' task required observers to detect a line of brighter illumination and shorter length. In this task the temporal properties of the stimulus (i.e., stimulus duration) were held constant. The 'temporal-only' task required a temporal discrimination in conjunction with luminance discrimination. Participants were required to respond to presentations of a line of brighter illumination and shorter duration. The spatial characteristics were held constant. In the third task ('combined') the temporal and spatial discriminations were combined. Thus, observers were required to detect a line of the shorter duration and shorter length. For all three tasks the stimulus presentation rate was 24 events per minute with six critical signals appearing per minute. The duration of each task was 7-minutes.

Pessimism and optimism were measured using the Optimism/Pessimism Inventory (Dember, Martin, Hummer, Howe, & Melton, 1989), a measure which treats these two traits as partially independent dimensions and provides separate scores for pessimism and optimism.

Self-reports of stress were measured using the Dundee Stress Questionnaire, a multi-dimensional instrument for assessing transient states associated with mood, arousal, and fatigue (i.e., pre and post task distress, engagement, and worry; Matthews et al., 1999). Hierarchical regression analysis was employed to examine the effects of pessimism and optimism on performance and stress for each task, as well as the

interaction between each of these traits and intermittent noise. Consistent with previous findings (Thropp, et al., 2003; Szalma, 2002), pessimism and optimism did not significantly impact overall performance, as measured by the proportion of correct detections and false alarms. In addition, neither of these traits interacted with noise in influencing performance or stress (p>.05 in each case). Contrary to the findings of Thropp et al., however, pessimism did not significantly impact stress state for any of the tasks in this study. Hence, the spatial task specific effects observed by Thropp et al. (2003), in which pessimism influenced task-engagement only in tasks with spatial uncertainty, did not extend to a case in which spatial discrimination is considered rather than spatial uncertainty.

However, optimism significantly predicted decreased task-engagement, after the pre-task state was accounted for in the model, $F(2,43)=11.69$, $p<.001$, $R^2=.35$, $\Delta R^2_{optimism}=.08$, but this main effect was restricted to the spatial only task. Optimism did not significantly predict post-task engagement in the temporal only or the combined tasks. Note that task was a within-subjects variable, so the individuals who reported decreased engagement after the spatial task did not report a similar change after the temporal task. In addition, preliminary analyses indicated that task order did not significantly impact post-task engagement.

Although the results of the current study did not confirm the hypotheses regarding joint resource capacities for spatial and temporal task components, results did indicate that dispositional optimism does impact one dimension of stress (task engagement). However, as observed in previous research, the effects of this variable are task dependent. Further, the direction the effect was unexpected, since in previous experiments increased

optimism predicted increased engagement rather than the decrease observed herein (Szalma, 2002; Thropp, et. al., 2003). However, these results do confirm that the dominance of spatial characteristics observed by Thropp et al. (2003), suggesting that the spatial dimension may be more salient than the temporal dimension. Future research will investigate the degree to which the apparent dominance of spatial characteristics over temporal characteristics generalizes to other task domains.

References

Dember, W. N., Martin, S. H., Hummer, M. K., Howe, S. R., & Melton, R. S. (1989). The measurement of optimism and pessimism. *Current psychology: Research and reviews,* **8***,* 102-119.

Hancock, P.A., Szalma, J.L., & Weaver (2002). The distortion of perceptual space-time under stress. *Proceedings of the 23rd Army Science Conference*, Orlando FL.

Hancock, P. A. & Warm, J. S. (1989) A dynamic model of stress and sustained attention. *Human Factors,* **31***,* 519-537.

Matthews, G. Joyner, L., Gilliland, K., Campbell, S. Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (vol 7), (pp. 335-350).

Szalma, J.L. (2002). Individual differences in the stress and workload of sustained attention. *Proceedings of the Human Factors and Ergonomics Society*, **46**, 1002-1006. CA: Human Factors and Ergonomics Society.

Thropp, J.E., Szalma, J.L., Ross, J.M., & Hancock, P.A. (2003). Individual differences in dispositional pessimism, stress, and coping as a function of task type. *Proceedings of the Human Factors and Ergonomics Society*, **47**, 1073-1077 . Santa Monica, CA: Human Factors and Ergonomics Society.



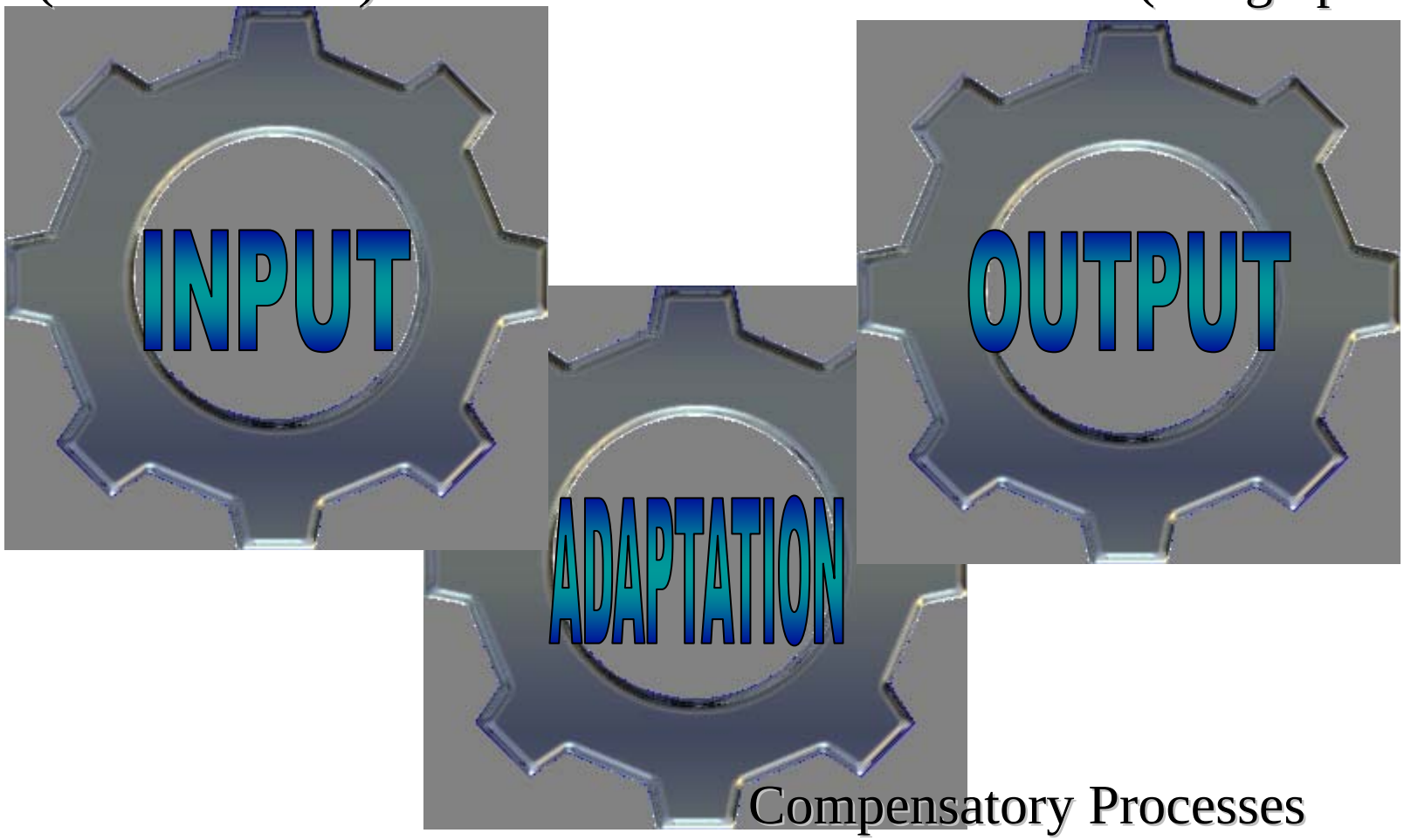
The Influence of Dispositional Optimism and Pessimism on Task Engagement for Spatial and Temporal Discrimination

J.L. Szalma, J.M. Ross, & P.A. Hancock
University of Central Florida

Trinity of Stress

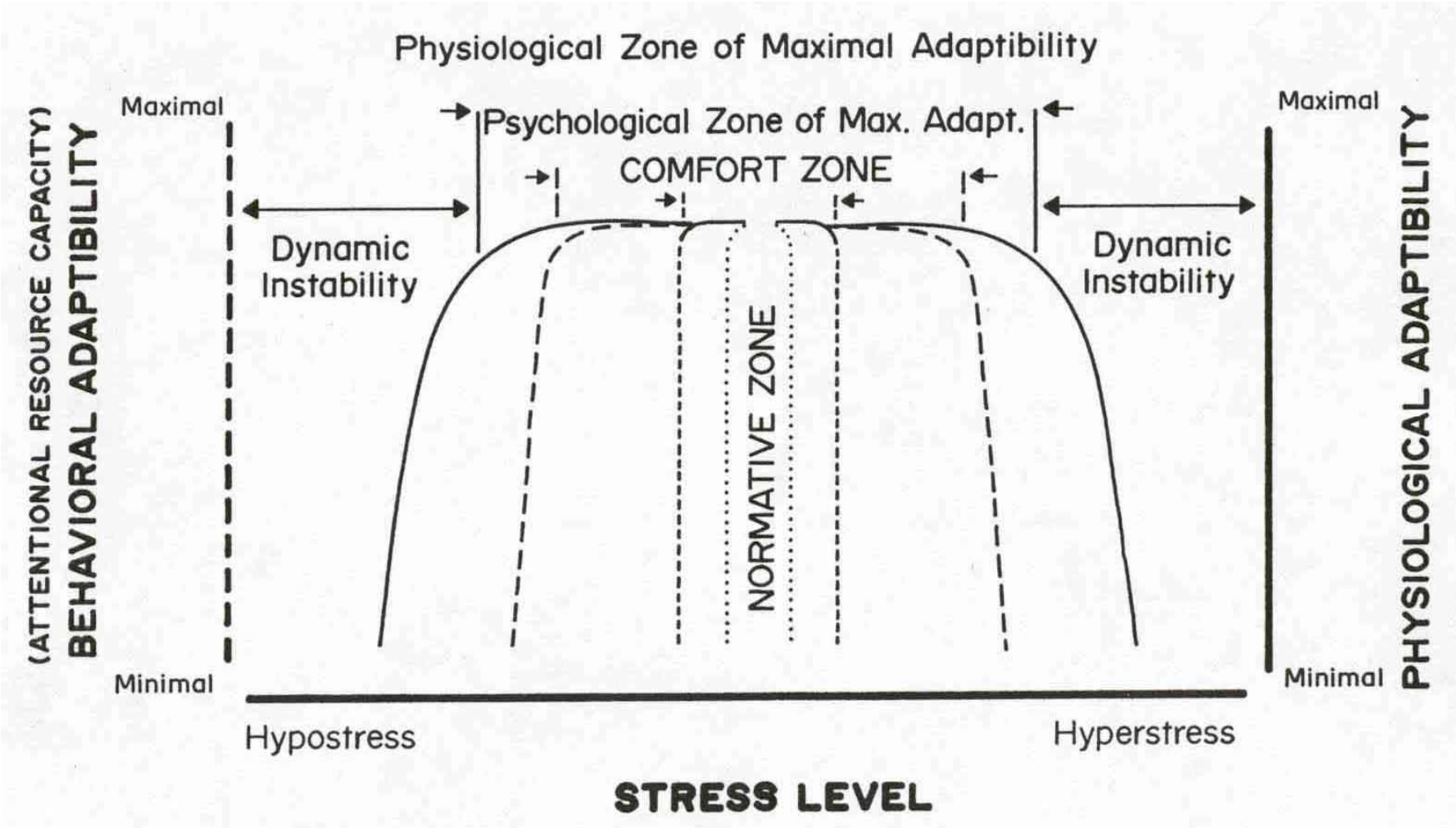
Stress Signature
(Deterministic)

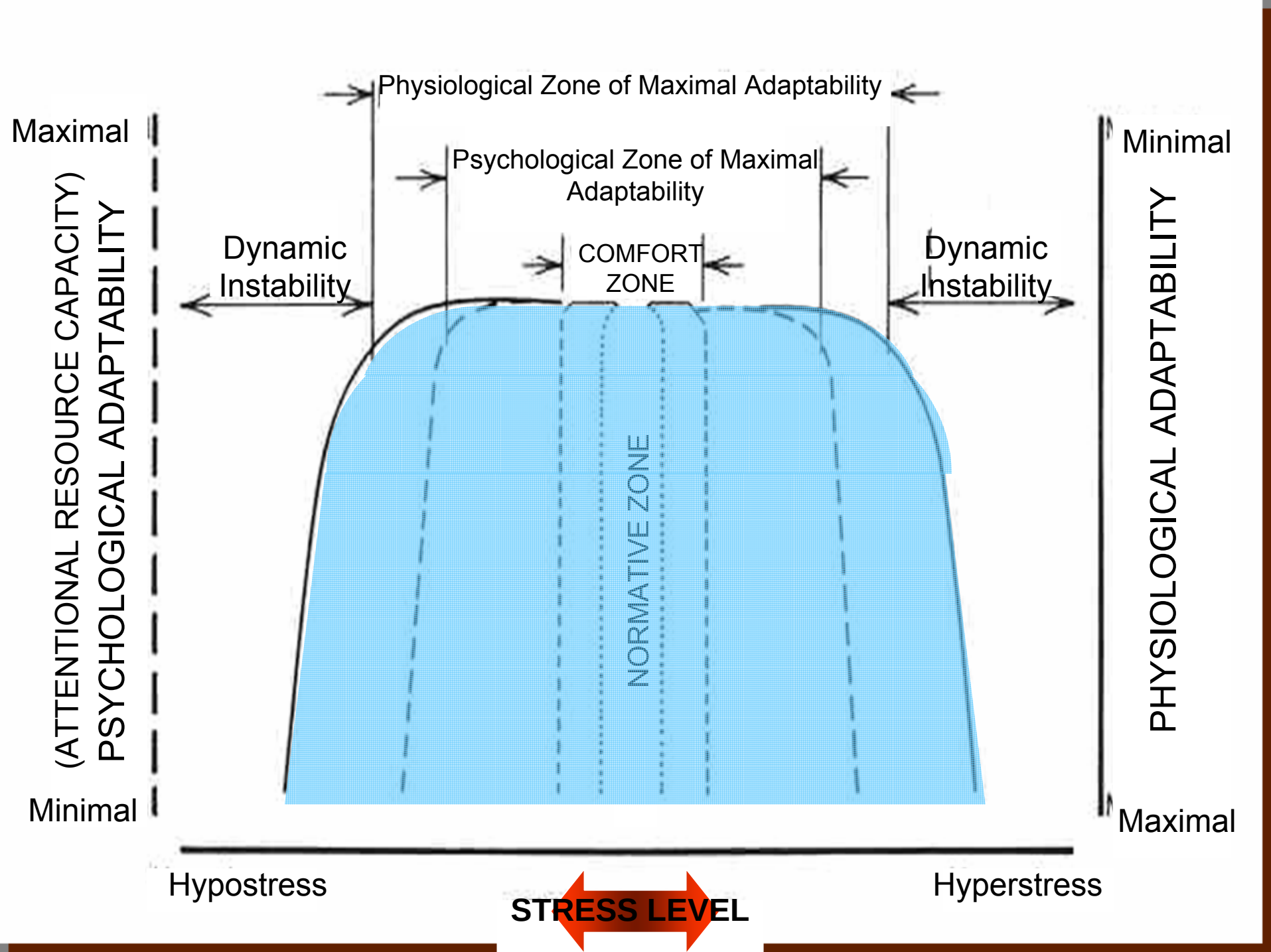
Goal-Directed Behavior
(Idiographic)



Compensatory Processes
(Nomothetic)

Hancock & Warm (1989)



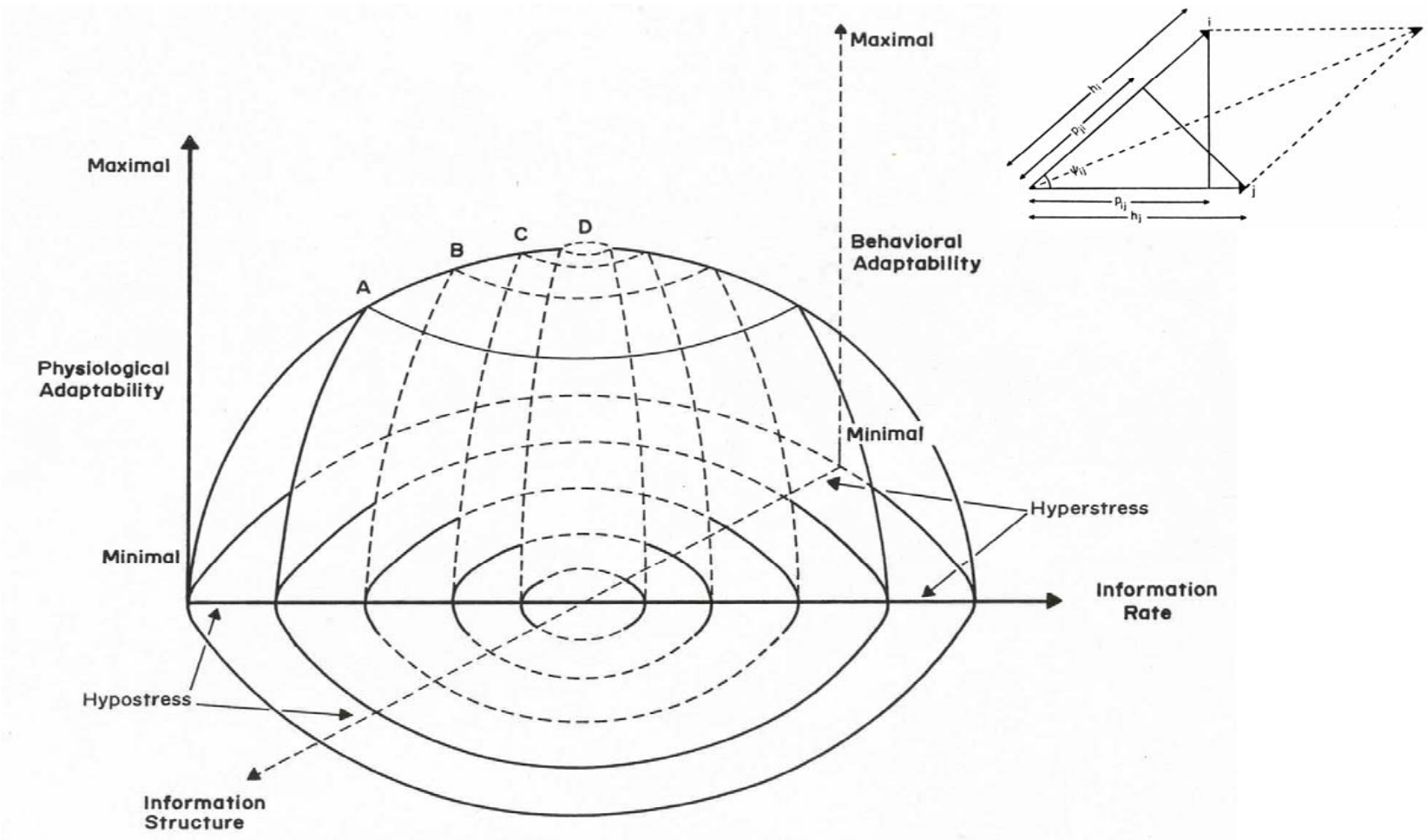

Physiological Zone of Maximal Adaptability
Psychological Zone of Maximal Adaptability
Dynamic Instability
COMFORT ZONE
Dynamic Instability
NORMATIVE ZONE
Maximal
(ATTENTIONAL RESOURCE CAPACITY)
PSYCHOLOGICAL ADAPTABILITY
Minimal
Minimal
PHYSIOLOGICAL ADAPTABILITY
Maximal
Hypostress
Hyperstress
STRESS LEVEL

Maximal Adaptability Model

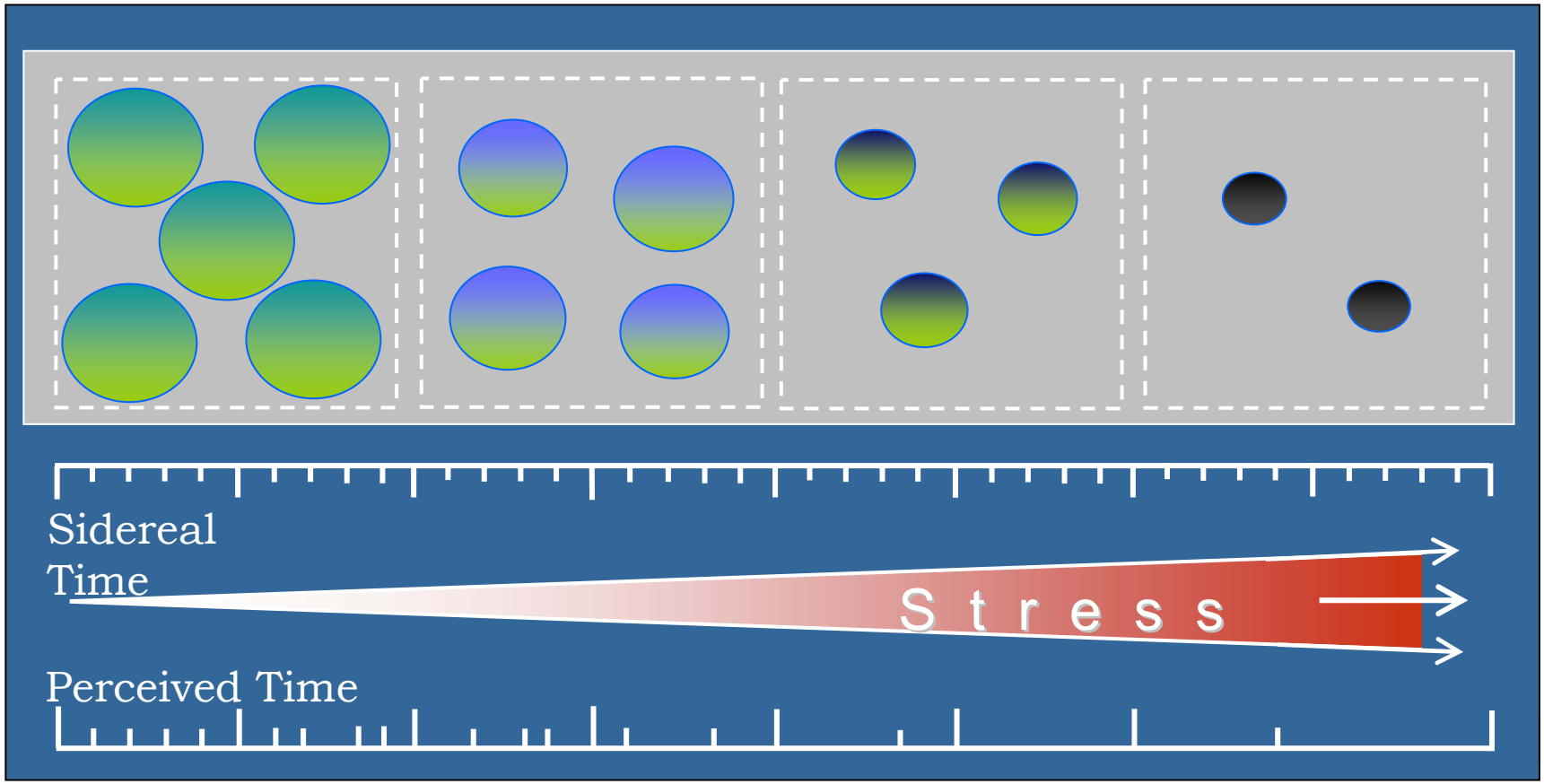
Hancock and Warm (1989)

- Recognition of task as a proximal stressor
- Two fundamental task dimensions.
  - Information structure – often expressed spatially.
  - Information rate – the temporal characteristics of the task.

Hancock & Warm (1989)



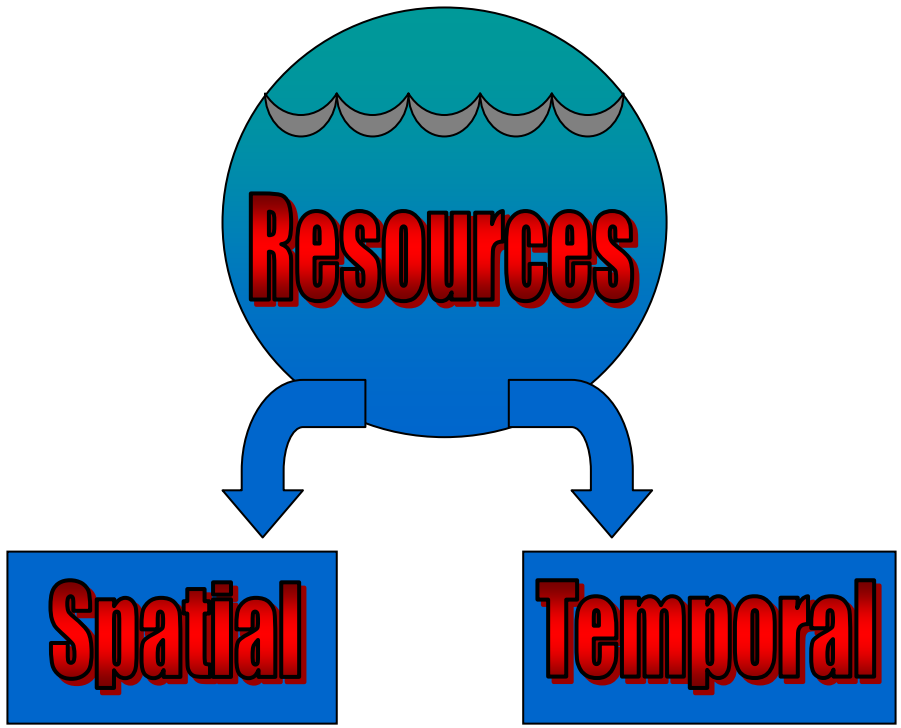
Distortion of Space-Time Under Stress



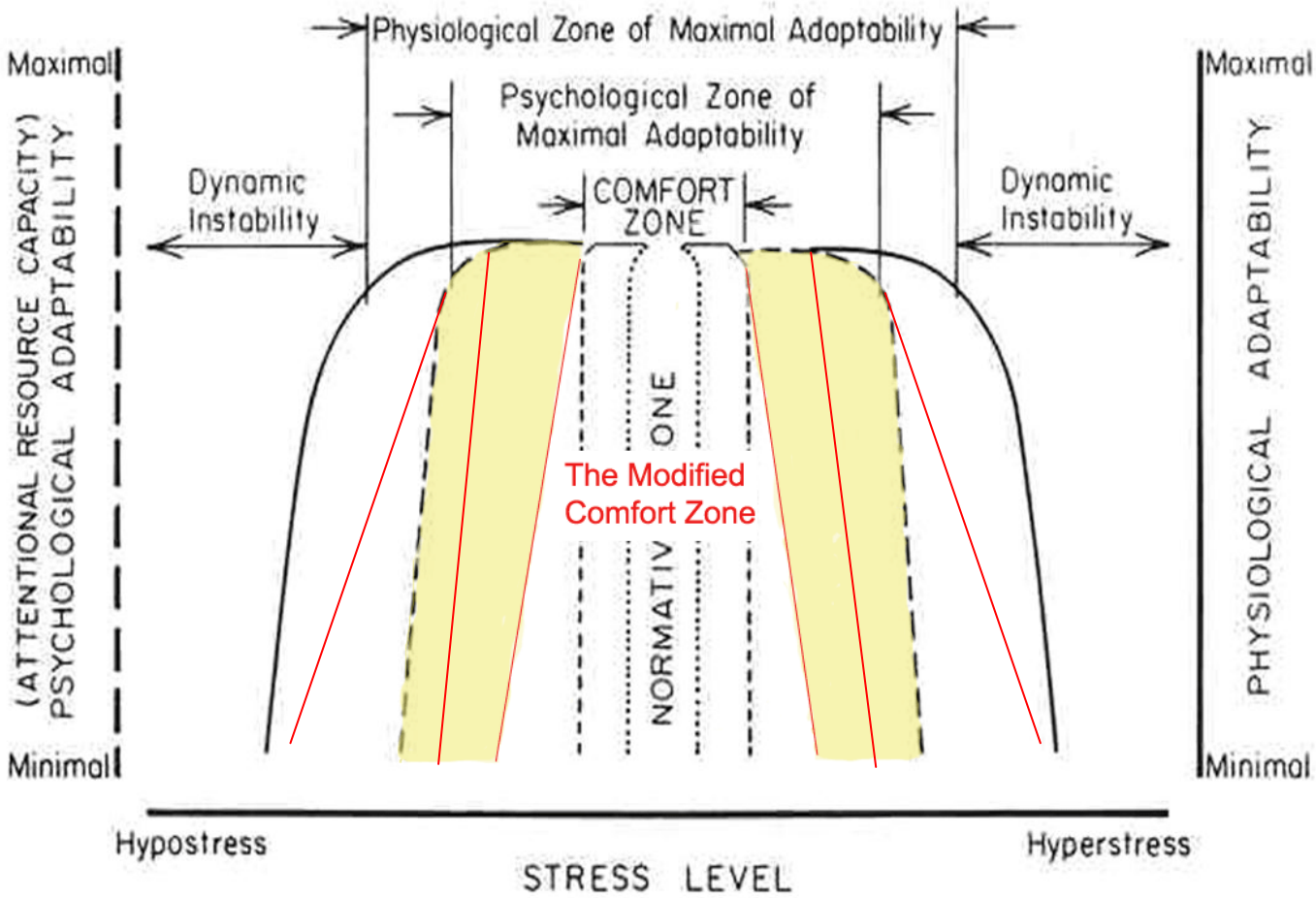
Hancock & Weaver (in press)

What is the Common Mechanism?

- One Possibility: Common Resource Capacity
- Narrowing occurs as a result of diminishing resources



Role of Individual Differences



Individual Differences in Stress Response

Individual's Traits

↓

Task Appraisal

↓

Stress Response



Goal-Directed Behavior
(Idiographic)

Dispositional Optimimsm/Pessimism

Performance Impairment in Pessimistic Swimmers (Seligman, Nolen-Hoeksema, Thornton, & Thornton, 1990)

Pessimistic candidates do more poorly in elections (Zullow, 1995)

Insurance agent productivity (Seligman & Schulman, 1986)

Pessimism and Attention

Performance impairment (sometimes; Helton, Dember, Warm, & Matthews, 1999)

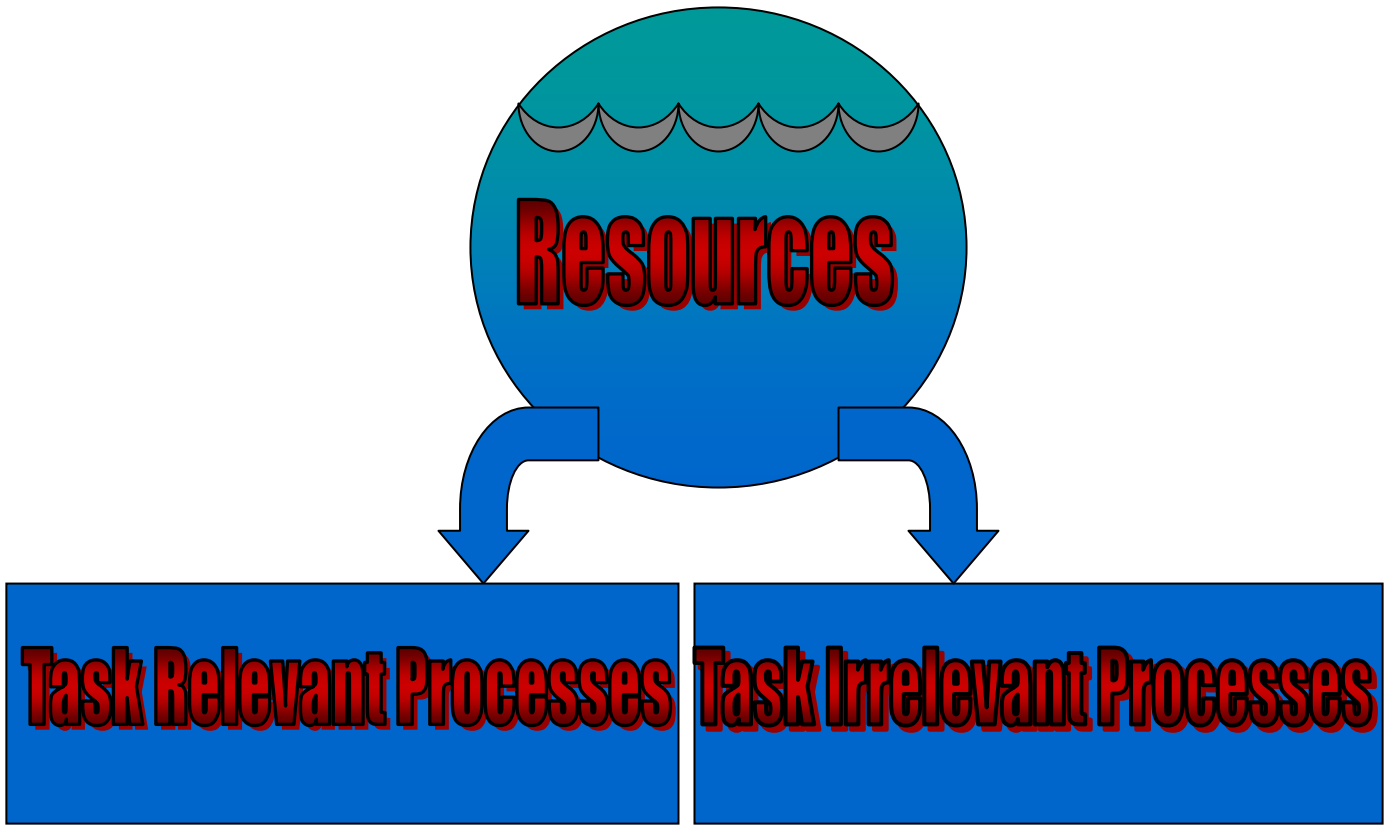
Increased Stress Symptoms (Helton et al., 1999; Szalma, 2002)

Maladaptive coping strategies (Scheier & Carver, 1987; Szalma, 2002)

Effects may depend on task characteristics (Thropp, Szalma, Ross & Hancock, 2003)

Relation of Pessimism to stress and performance in attention is generally stronger than that associated with optimism (when measured as distinct constructs)

Traits and Resource Sharing



Hypotheses

1. Individuals high in pessimism (low on optimism) would exhibit greater stress symptoms than those low in pessimism (high in optimism)

2. These effects should be greater under more demanding task conditions: combination of spatial and temporal characteristics

3. Trait effects should be greater when an external stressor is applied (white noise)

Experimental Procedure

- Participants
  - 46 Undergraduates (23 Males, 23 Females)
- 2 (noise) by 3 (task) mixed design with repeated measures on the second factor.

Experimental Procedure

- Noise: 85 dBA intermittent white noise
- Conjunctive DiscriminationTasks:
  - *Spatial Dominant*: Spatial + Luminance discrimination
  - *Temporal Dominant*: Temporal + Luminance discrimination
  - *Combined*: Spatial + Temporal discrimination

Questionnaires and Procedure



- **<u>Optimism/Pessimism Inventory (OPI)</u>**:
  Pessimism/Optimism measured as partially independent constructs (Dember, Martin, Hummer, Howe, & Melton, 1989)

  <u>Pre-DSSQ</u>:

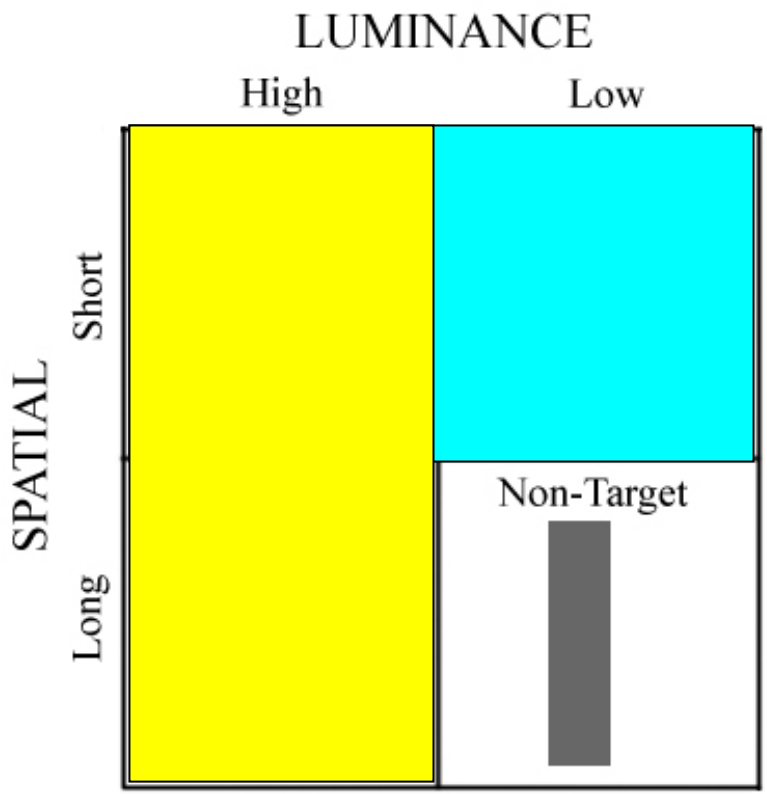
  -Pre-Task Engagement, Pre-Task Worry, Pre-Task Distress

  <u>Post-DSSQ</u>:

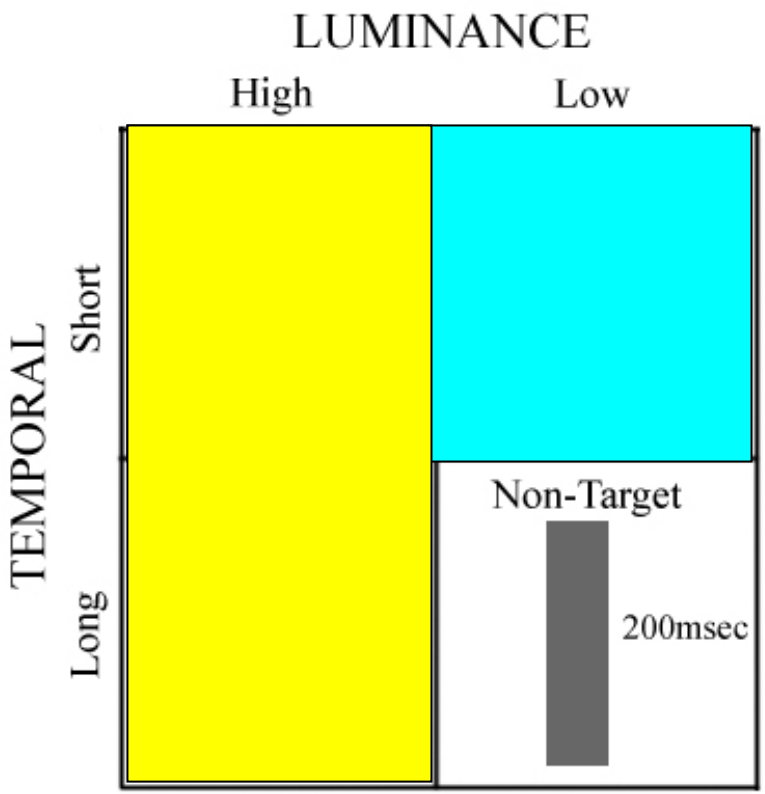
  -Post-task Engagement, Post-task Worry, Post-task Distress

  -Task-focused, Emotion-focused, Avoidant Coping

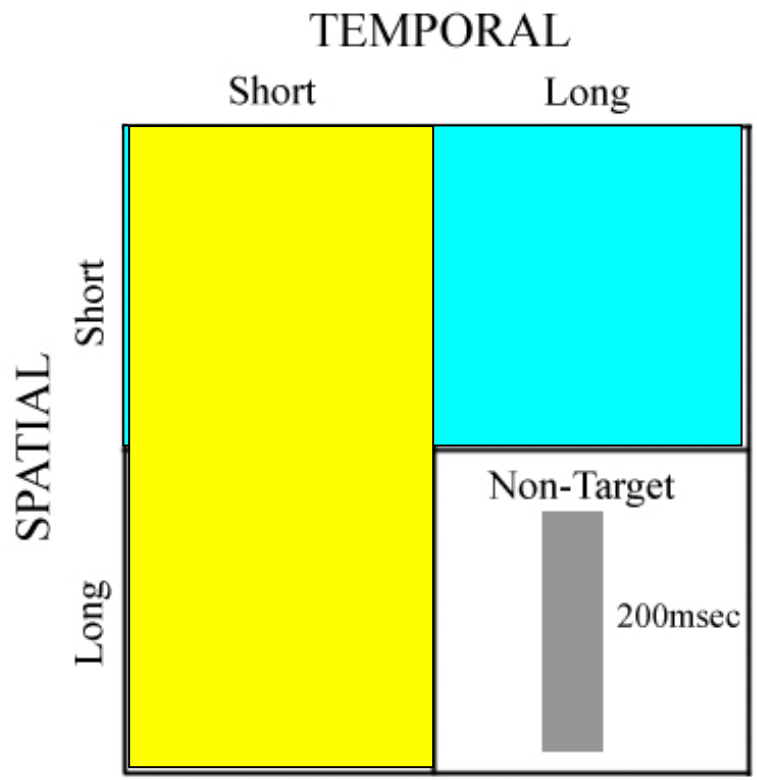
Spatial-Dominant Task Stimulus



Temporal-Dominant Task Stimulus



Combined Task Stimulus



Results: Performance

- Optimism and pessimism did not predict performance (Hits, FA, d’, c)

- Neither trait interacted with noise to influence performance or stress

Results: Pessimism and Stress

- Pessimism predicted greater post-task Distress, but only in the spatial-dominant condition (R^2 = .11, p<.05)

- This effect was not significant when the pre-state was entered first into the regression (ΔR^2 = .01, p>.05)

Optimism and Distress

- Optimism predicted less post-task Distress in all three tasks (R^2 = .08 --.11, p<.05 in each case)

- These effects were not significant when the pre-state was entered first into the regression (ΔR^2 < .01, p>.05)

Optimism and Task Engagement: Spatial-Dominant Task

- Optimism did not predict changes in post Task Engagement in tasks with a spatial component, (R^2 <.02, *p*>.05 in each case)

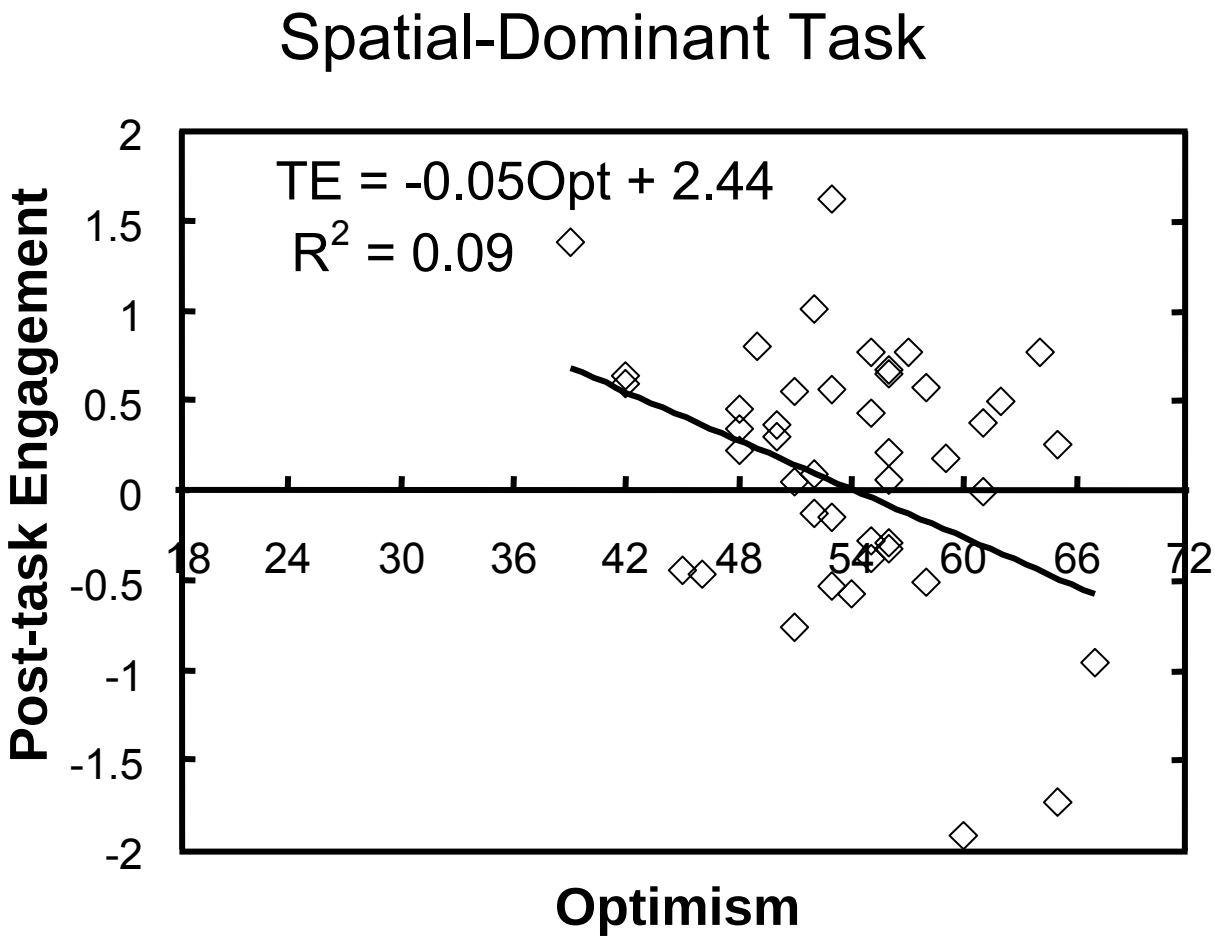
- In the spatial task, optimism predicted greater post Task Engagement after the pre-state was entered first (R^2 =.35; ΔR^2 = .08, *p*<.05)

Optimism and Task Engagement: Temporal-Dominant Task

- Optimism predicted decreased post Task Engagement in the temporal-dominant task (R^2 =.09, p<.05)

- This effect was not significant when the pre-task state was entered first (ΔR^2 <.01, p>.05)

Post-Task Engagement as a Function of Optimism



Optimism and Worry

- Optimism predicted increased post task Worry, but only in the temporal-dominant task (R^2 =.1, p<.05)

- This effect was not significant when the pre-task state was entered first (ΔR^2 =.04, p=.07)

Conclusions

- Optimism may exert a greater influence on stress response than previous experiments indicated
- This effect is task dependent
- Effects vary across dimensions of stress-state
- White noise did not interact with either trait
- Effects of optimism/pessimism on stress state was not exacerbated by the combination of spatial/temporal demands
- The stress-trait relation varied depending on task dimension emphasized (spatial, temporal)

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, Dr. Elmar Schmeisser, Technical Monitor (Grant# DAAD19-01-1-0621). The research was facilitated by a DARPA-funded program under Grant NBCH1030012, CMDR Dylan Schmorrow, Technical Monitor. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government.

Analysis of Distributed Team Communications:

Team Performance and Training Implications

Guthrie, Jr., J.W., Asberg, K., Bower, C.A.,

Salas, E. & Jentsch, F, University of Central Florida

The nature of teamwork is complex and includes aspects such as mutual performance monitoring, backup behavior, intrateam feedback and shared belief among team members that team success is dependent on their interaction (Salas, Bowers, & Cannon-Bowers, 1995). Among the inherent characteristics of teamwork are team communication and coordination. Both concepts have been studied and a relationship between these behaviors and team performance has been demonstrated. However, this relationship is quite complex and is likely to depend on the specific nature of the task, the type of team interdependency used, and general team type.

Although a link between successful performance and particular types of communication can (sometimes) be established among teams who perform the same task under the same condition, little is known about differences in communication when conditions are manipulated and teams are randomly assigned to different groups. For example, few studies, if any, have investigated differences in communication patterns, regardless of performance, between teams under high stress compared to teams in a low stress condition. In the sample used for this particular study (N=8), initial results indicated that high stress teams experienced a significantly higher level of perceived stress when compared to teams in the no stress condition (p=.001); however, no significant group differences in actual outcome (team performance) were found (p=.9j5). Hence, the next step in the investigation was to examine what process factors would allow high stress teams, despite experiencing more time pressure, to perform at the same level as

their low stress counterparts. The process in this case refers to differences in patterns of communication among team members in the different conditions during the task.

Method

Participants

Undergraduate students at a large southeastern university were used for the present study. The students who volunteered were randomly assigned to either the navigator ("on-site/computer) position or interviewer ("headquarter"/no computer) position, forming 8 two-person distributed teams.

Apparatus

A COTS first person view computer game, Operation Flashpoint, was employed. For the purpose of this study, the specific training scenario and the timed task scenario were developed using the mission editor function of Operation Flashpoint to create a simulated task for a reconnaissance team with reciprocal interdependence.

Procedure

The practical component of the experiment was comprised of two phases: one training phase (15 minutes) and one performance phase (30 minutes). The training phase was designed to familiarize participants with the computer game and the materials to be used when completing the mission. The session included instructions regarding the sheet of statements to be verified. The training phase also included general directions, and allowed for participants to ask questions.

Teams were randomly assigned to either the experimental condition (time pressure) or the control condition (no time pressure). The time pressure condition consisted of the experimenter explicitly stating the time limits of the task (30 minutes) following the training, and giving

information about time remaining (e.g., "7 minutes left") at two-minute intervals. Furthermore, using the same two-minute intervals, false performance feedback (N+2) was given to induce additional time-pressure. False feedback involved the receipt of information from the experimenter regarding the number of verifications an average team would have at the particular point of the task.

Results

The first analysis conducted was a 2 (high stress v. low stress) x 6 (antecedent) x 6 (consequence) cross-tabular analysis to determine if any specific patterns of communication emerged between the two types of teams. This analysis revealed that low stress teams were more likely than high stress teams to use a Planning → Acknowledgement communication sequence, while high stress teams were more likely than low stress teams to use Action/Command → Acknowledgement communication sequence. One interpretation of this group difference could be related to the team process as previously discussed. Although acknowledging the receipt of a plan or command does not imply that teams actually carry out what has been suggested, it may be that high stress teams attempt to save time by using brief commands rather than non-specific plans to communicate ideas. In order to determine whether or not the processes of high stress teams differ from those of low stress teams, further analyses of 2, 3, and 4 communication sequences were conducted, but did not produce any meaningful statistically significant results.

While the sequential analyses did not add any additional information as to statistically significant differences in team processes, we conducted a series of cross-tabulations to determine if any trends in communication were found between the two groups. From these analyses we wanted to determine, 1) do low stress teams have more closed communication sequences and 2) if there was a difference in the type of closing statement used (Acknowledgement or Response).

The results of these analyses show that low stress teams did produce a higher percentage of closed sequences than high stress teams (Figure 1).

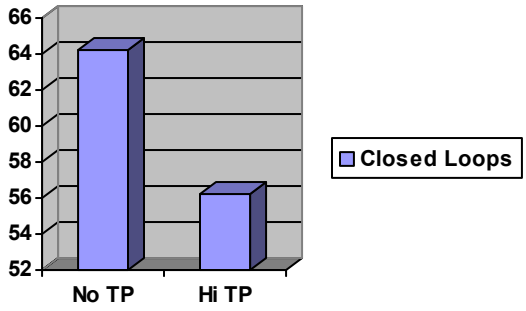


Figure 1: Percentage of closed loop communications by condition

In addition, high stress teams used acknowledgements to close the loops more than the low stress teams, while the low stress teams were more likely than the high stress teams to use responses to close the sequence (Figure 2).

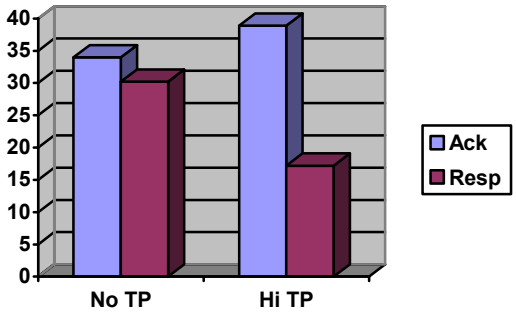


Figure 2: Percentage of overall communications closed by acknowledgements and response

Discussion

Although teams who were under high stress did not perform any differently, this was to be expected since teams have typically been found to be adept at handling or absorbing stress. However, there were differences in how the teams communicated based on their condition.

Teams not under stress had a higher percentage of closed loops, indicating more complete communication patterns. This finding was expected based on the finding of Bowers et al. (1998). In addition teams not under stress used responses more often to close those loops, indicating the exchange of more information. High stress teams closed loops with acknowledgements or 1-bit statements more often, possibly in an effort to save time. The use of acknowledgements instead of responses to close communication loops and the use of Action/Command→Acknowledgment communication sequences indicates a difference in communication processes that can, in part, explain why high stress teams did not perform any differently than low stress teams.

While none of the sequential analyses produced any statistically significant differences, the trends found in the cross tabulations more than likely would have been statistically significant if more teams were included in the study. In addition to replicating the results from this study, future research should investigate other types of teams to determine if the same or similar communication patterns exist among different types of teams or if these team processes are team specific. In addition, further exploration into the benefits of using sequential communication analysis in the study of team process is warranted.

References

Bowers, C.A., & Jentsch, F., Salas, E., Brown, C.C. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors, 40(4),* 672-679.

Foushee, H.C., & Manos, K. (1981). Information transfer within the cock-pit: Problems in intracockpit communications. In C.E. Billings & E.S. Cheaney (Eds.), *Information transfer problems in the aviation system* (Report No. NASA TP-1875). Moffett Field, CA; NASA-Ames Research Center.

Jentsch, F., Sellin-Wolters, S., Bowers, C.A., & Salas, E. (1995). Crew coordination behaviors as predictors of problem detection and decision making times. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1350-1253). Santa Monica, CA: Human Factors and Ergonomics Society.

Mosier, K.L., & Chidester, T.R. (1991). Situation assessment and situation awareness in a team setting. In Y. Queinnec & F. Daniellou (Eds.), *Designing for everyone: Proceedings of the 11th Congress of the International Ergonomics Association* (pp. 798-800). London: Taylor & Francis.

Orasanu, J.M., & Fischer, U. (1991). Information transfer and shared mental models of decision making. In R.S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 272-277). Columbus: Ohio State University.

Salas, E., Bowers, C., & Cannon-Bowers, J. (1995). Military team research: 10 years of progress. *Military Psychology*, *7(2),* 55-75.

Thornton, R.C. (1992). *The effects of automation and task difficulty on crew coordination, workload, and performance.* Unpublished doctoral dissertation, Old Dominion University, Norfolk, VA.

Submitted for presentation at 2004 APA Annual Meeting, July 28-31, Honolulu, HI

Sequential Analysis of Distributed Team Communications: Implications for Team Performance and Training

Kia Asberg, Joseph Guthrie, Clint Bowers, Eduardo Salas & Florian Jentsch

Abstract

The present study assessed the role of sequential communication in team performance. Communications among undergraduate students who participated in a complex task as part of a distributed team were investigated. Results indicated a significant group difference in perceived stress (time pressure) between groups in the high and low stress (control) condition, but no significant group difference in performance was found. In contrast, significant group differences in communication patterns was found, indicating that high stress teams used more one-bit statements to close communication loops.

Introduction

- Sequence analysis of team communication may reveal more about effective teams than what is generated by mere frequency counts (Bowers, Jenstch, Salas, & Braun, 1998).
- Elements of sequential data analysis may be useful tools for understanding team communication and for identifying team training needs (Sanderson & Benda, 1998).
- Further evidence of the notion that with communication "more is not always better" was found by Thornton (1992).
- Little is known about differences in communication when conditions are manipulated and teams are randomly assigned to different groups

Research Questions

Does stress change the communication patterns found in Bowers et al. (1998)?

- Do high stress teams have different communication frequencies and sequences than low stress teams?

Method

Participants

- 16 undergraduate students at a large southeastern university were assigned to 8 two-person teams
- Team members were randomly assigned to either the navigator or interviewer position

Apparatus

- A COTS first person view computer game, Operation Flashpoint, was employed
- Enemies and ammunition were removed
- Stationary objects to be located by the team (such as computers, radios, civilians, and vehicles) were inserted into the virtual environment

Procedure

- 2 (navigator, interviewer) x 2 (no time pressure, time pressure) design was used
- Participants were given one 15 minute training phase and one 30 minute performance mission
- Participants were required to work together with one team member navigating the virtual environment and one team member directing and reading statements, to verify a list of statements
- False feedback was used to increase time pressure

Sequential Analysis Communication Categories

Category	Defined as:
Plan	A statement that communicates an expected or anticipated plan, event, goal, etc.
Action/Command	A statement/command that outlines an action (e.g., gives directions)
Acknowledgement	A one-bit statement that confirms the receipt of a command or fact
Response	A statement in response to a plan, action, uncertainty
Uncertainty	A statement of uncertainty expressing a question that seeks information from the other team member
Facts	A statement about the environment or other fact

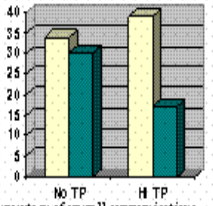
Bowers et al., 1998

Communication Category by Condition

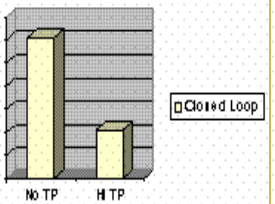
		Time Pressure (TP) Condition		TOTAL
		No TP	Hi TP	
Communication Categories	Plan	26	23	49
	Action	227	323	550
	Ack	531	674	1205
	Response	343	259	602
	Uncert	609	642	1853
	Facts	178	264	442
	Total	1914	2185	4099



Results: Communication Category Sequences by Time Pressure Conditions



Percentage of overall communications closed by acknowledgements and response



Percentage of closed loop communications by condition

Conclusion

Although teams who were under high stress did not perform any differently, this was to be expected since teams have typically been found to be adept at handling or absorbing stress. However, there were differences in how the teams communicated based on their condition. Teams not under stress had a higher percentage of closed loops, indicating more complete communication patterns, and used responses more often to close those loops, indicating the exchange of more information. High stress teams closed loops with acknowledgements or 1-bit statements more often, possibly in an effort to save time.

MILITARY PSYCHOLOGY, 2004, *16*(1), 135–147
Copyright © 2004, Lawrence Erlbaum Associates, Inc.

Motivational Effects of Adding Context Relevant Stress in PC-Based Game Training

Christina S. Morris, Ph.D.
Laboratory for Applied Sensory Engagement Research and Department of Psychology
University of Central Florida

Peter A. Hancock, Ph.D.
Department of Psychology and Institute for Simulation and Training
University of Central Florida

Ed C. Shirkey, Ph.D.
Department of Psychology
University of Central Florida

This work was designed to examine the effects of contextually relevant stress on personal computer (PC)-based game training. Off-the-shelf PC-based games are being applied to many training situations because of their affordability, flexibility, and teaming capabilities. The ultimate purpose of training is to transfer superior performance to the real world. In this respect, 1 of the major drawbacks to using games as training tools, especially for military applications, is the absence of the surrounding context. In response to this omission, we examined the effects of adding context-relevant stress to infantry game-based training by exposing 1 group of participants to a graphically intense and stressful experience while the control group viewed an unstressful analog. Pre–post self-reported stress levels confirmed the efficacy of this manipulation. The stress condition produced significantly higher scores on "mission success"; however, no differences were evident in participants' use of trained tactics or game functions. Supplementing context-relevant stress in game training shows promise for enhancing individuals' motivation to succeed.

Requests for reprints should be sent to C.S. Morris, Department of Psychology, University of Central Florida, Orlando, FL 32816. E-mail: cmorris@mail.ucf.edu

In the early 1930s, the U.S. Army Air Corps experienced a number of aviation fatalities due to encounters with hazardous weather conditions. The deaths were primarily the result of the pilots not having been trained to fly an aircraft using instruments alone. Safe operations were dependent on the occurrence of clear and visible flying conditions. With the need for round-the-clock operations, the military recognized the necessity for simulators to train operators for such hazardous conditions. Fortunately, Link had already developed the first simulator for instrument-based flight training in 1929 and was able to rapidly build more simulators for this purpose. This transition from the traditional classroom setting to simulated environments provoked a more rigorous examination of the field of training, and traditional instructional strategies were then criticized for merely providing training without practical application to real world contexts (see Rogers, 1969). Simulation-based training represented a promising response to this criticism, and as such simulators have been employed in military and commercial agencies worldwide.

Historically, training programs have had difficulty exposing individuals to unexpected stimuli and sources of task demands that are typically experienced in the real world counterpart. With recent advances in technology, however, the delivery of training in more realistic and high-stress settings has become feasible. The replication of the real world setting and the degree of operator stress imposed is suggested to be the driving component for effective skill acquisition and training transfer (e.g., Bowen, 1987). For this reason, the features employed in simulators (e.g, graphics, sound systems, haptics) are continuously being redesigned and improved for the replication of real world environmental conditions (Hancock, 2003).

Simulation-based training can only be feasible and cost effective when there is a high degree of "transfer" of the trained skills from the simulator to the real world. However, the ability of an individual to retain and transfer the skill to real world settings hinges predominantly on their "willing suspension of disbelief" and their degree of acceptance of recreated reality. The effective human state of immersion within the environment varies considerably with individual human's ability, desire, and motivation to temporarily suspend that disbelief (Brown, 1999). Being a subjective capacity, such achievement then is not necessarily contingent on the reality produced by the simulation system itself. The effectiveness of the duplication of the physical target environment using advanced simulator features, then, is relative, varying among individuals. Although advanced technologies may deliver realistic training scenarios in a simulated environment, they may not improve training (Kantowitz, 1988). Rather, the degree of a human's immersion within that scenario and their resultant level of emotional and behavioral reactivity are suggested to primarily generate effective training. These necessary human qualities that foster optimal immersion are found consistently in studies employing a "gaming" condition. When the elements of a game are present, part of the physical fidelity or reproduced realness of a simulated environment may be sacrificed while immersion itself re-

mains still at an optimal level for training effectiveness. Thus, personal computer (PC)-based gaming environments can be highly effective training tools.

PC-BASED GAME TRAINING

Although the effectiveness of replicating the real world in a fully simulated training environment has some advantages, there are also many unique benefits of training that utilize PCs. Among these are the portability, convenience of use, and the flexibility to quickly evolve and adapt to software improvements that become available. The benefits of interlinked PCs as well as their remote site capabilities guarantee PC-based systems to be the primary vehicle for advanced distributed learning environments. The most recent "off-the-shelf" (OTS) PC-based games have also immediately responded to distributed needs by providing remote multiplayer capabilities through networks. These low-cost and networked gaming environments have been shown to be effective for training shared cognition, team skills (Baker, Prince, Shrestha, Oser, & Salas, 1993), situational awareness (Homan, 1998), enhancing interpretation of explicit and implicit information, inductive reasoning, metacognitive analysis, and problem solving (Ricci, Salas, & Cannon-Bowers, 1996). It is therefore unsurprising that most, if not all, military agencies are exploring the advantages of low-cost, OTS, PC-based games for their training requirements (Morris & Tarr, 2002). Traditionally, OTS PC-based games have been solely created and intended for entertainment purposes, although most recent developers of infantry games are designing their products according to reputable military data sources (Coleman & Johnston, 1999). The benefits of these efforts for doctrinally correct infantry games are twofold. First is the higher face validity for using their games in military training. Second is the increase in public knowledge, interest, respect, and awareness of military operations, especially for potential recruits and reservists.

Set against the aforementioned advantages of using games for training are some intrinsic drawbacks, one of which is the topic of this investigation. A major insight in behavioral research in the last decade of the 20th century concerns the importance of *context* in predicting human behavioral response (see Flach, Hancock, Caird, & Vicente, 1995). It has become progressively more clear that results from the sterile laboratory have only restricted applicability in predicting performance in the real world. The complicated tapestry of environmental influences are not merely "unwanted" variance to be isolated and excused from the experimental procedures. Rather, they are crucial interactive facets of the performance portraiture.

Research has shown that when PC-based training involves warlike activity as its objective, the context that the user experiences can determine the degree of stress provoked (Wilson, Skelly, & Purvis, 1999). For military contexts, the presenting

problem in using PC-based games for training is the transformation from the objectives, strategies, and goals inherent in entertainment to those of the intensely stressful military settings while capturing the motivational qualities of the games. Infantry subject matter experts (SMEs) have reviewed doctrinal and contextual correctness of emerging OTS infantry games and concluded that games cannot provide the level of realism needed to prepare individuals for tactics in the midst of real war (Tarr, Morris, & Singer, 2002). Most criticisms of using PC-based games for training, especially in the military environment, have argued that the behavioral models for the agents are unrealistic and the context is amusement rather than the preparation for stressful conflict. Essentially all crucial military tasks possess or occur in conjunction with a high degree of potentially intense and unpredictable stress, and training for these tasks mandates that advanced simulations produce similar conditions, generating similar stresses, as the real world task counterpart. Contextually correct simulations generate higher levels of stress during training, thereby reducing the user's level of stress in actual operational performance (see Driskell & Salas, 1991). Difficulties in games and simulations also arise when the real world environment is unpredictable, as is virtually always the case with stressful contexts such as combat (Driskell, Johnston, & Salas, 2001). When designing stress training, it is therefore crucial to design the stress conditions based on expected real world conditions (Johnston & Cannon-Bowers, 1996). However, stress is extremely difficult to operationally define, as evidenced in the variety of efforts across the literature (see Hancock & Desmond, 2001; Hancock & Warm, 1989). Accordingly, distinguishing how to apply the correct contextual elements of stress and naturally unpredictable tasks to games and simulations requires major empirical efforts. In consequence of these concerns, the purpose of this study was to determine if the effectiveness of PC-based game training could be improved by applying a supplementary stressful context. We hypothesized that the experience of stress will positively influence outcome performance. Specifically, we predicted that individuals who were exposed to graphically intense and technologically advanced realistic war scenes just prior to PC game-based training would perform better than a control group on the posttraining game-based task.

EXPERIMENTAL METHOD

Experimental Participants

Twenty-four undergraduate and graduate student volunteers (15 males and 9 females) participated in the study, with a mean age of 22 and no prior experience in the military. Of the individuals who were exposed to the stressful film condition, 5 reported having seen the film prior to the study. Of the individuals exposed to the con-

trol film condition, none reported having seen the film previously. Both groups were statistically equivalent on self-reported gaming habits and gaming experience.

Experimental Materials

An informed consent form, debriefing, and participant biographical questionnaire, which included gaming habits, were specifically designed for this project. A 15-min segment of the video *Saving Private Ryan* was used for the context enhanced group, while a 15-min segment of a black and white *WWII–Invasion of Normandy* documentary was used for the control comparison. Both film segments portrayed their respective versions of the beach invasion by US infantry. The Activation-Deactivation Checklist (ADC; Thayer, 1967) and the Emotional Assessment Scale (EAS; Carlson, Collins, Stewart, & Porzelius, 1989) were used to indicate changes in participants' pre and post stress, mood, and arousal levels. Methods used for the infantry training were developed by an SME, who also selected the PC-based game environment most optimal for exercising infantry operations and tactics. From the methods, training protocol and supplementary materials were prepared for a train-up session of infantry tactics such as cover and concealment. The arctic mission, novice level, in the Delta Force OTS game was used for the train-up involving performance goals, tactics, and game use. This same scenario was used as the task environment where performance was measured. The performance score sheet was developed by the SME for use by a human observer who was blind to the purpose of the study, but who had high expertise in gaming and who was trained on the infantry procedures, tactics, and protocols used during train-up.

Experimental Procedure

The participants arrived at the test facility where they completed the informed consent, the biographical questionnaire, the ADC, and the EAS and were randomly assigned to one of two training conditions. Those in the experimental group were asked to watch the 15 min of graphically intense war scenes from the beach invasion portion of the movie *Saving Private Ryan*. The participants in the control group viewed an alternative nonstimulating black and white war clip of the beach invasion. Immediately following the 15-min viewing, the participants completed the post ADC and EAS and were then trained for another 15 min on infantry tactics, weapon use, and mission objectives using the Delta Force environment. They were then asked to meet the mission objectives and reach waypoints in the mission, without their player being injured, but with the ability to restart the mission (when deaths occurred) as often as needed for a period of 20 min, during which they were observed and scored on mission success, game adaptation, and use of learned tac-

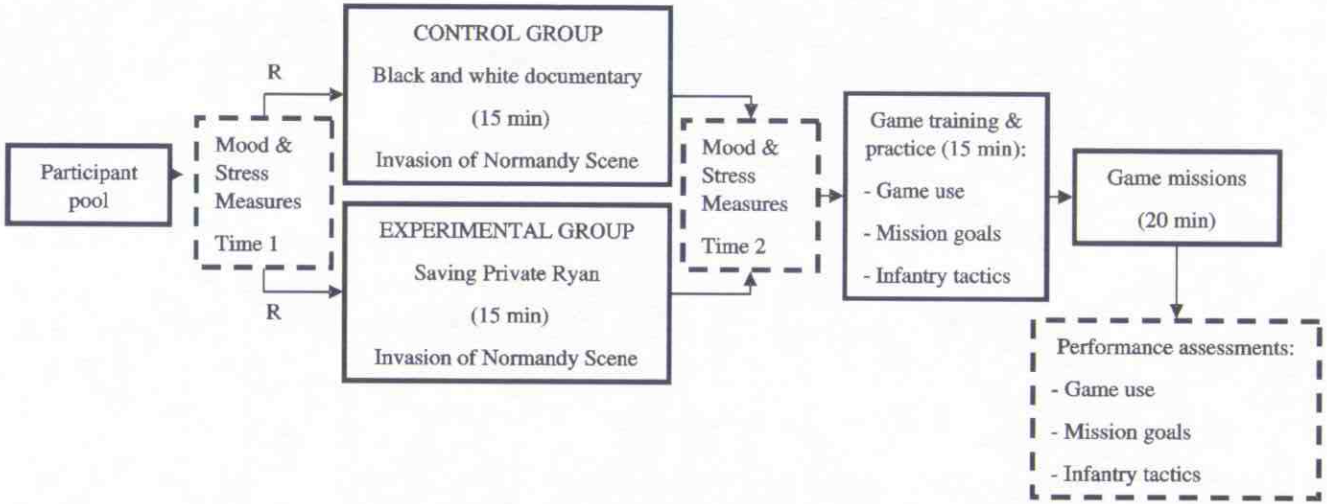


FIGURE 1 Experimental design.

tics by an game expert who was knowledgeable about infantry tactics and blind to the purpose of the study.

RESULTS

Between-Group Comparisons

Prior to the video exposure, independent sample *t*-tests with an alpha of .05 revealed no initial pretest differences between the experimental and control group on any of the ADC and EAC self-report indicators of stress. Following the 15-min video exposure, the experimental group had significantly higher mean scores than the control group on self-reports of "feeling clenched," "intensity," "anxiety," and "disgust." Individuals who were in the experimental group and had seen the film before did not differ significantly from those who had not seen the film on any of the ADC or EAC indicators of stress (see Table 1).

Between-group multivariate analysis of variance was performed on three observed dependent variables: mission success, game adaptation, and use of learned tactics, with condition assignment (experimental or control) as the independent variable. With the use of Wilk's criterion, the combined dependant variables (DV) were significantly affected by condition assignment, $F(4, 19) = 7.15, p < .001$. The results reflected a strong association between condition assignment and the combined DVs, partial $\eta^2 = .60$.

Univariate analysis revealed that those in the context enhanced condition scored significantly higher on "mission success" (*M* adjusted success score = 16.38) than those in the control group (*M* adjusted success score = 12.66), univariate $F(1, 22) = 6.075, p < .025$. However, there were no differences between the groups on "game adaptation" scores ($p = .415$) or "use of learned tactics" ($p = .787$; see Table 2). We

TABLE 1
Means, Standard Deviations, Independent Samples T, and P Values for Post Manipulation Stress Indicators in the Context Enhanced and Control Group

	Control Group		*CE Group*			
Stress Indicators	*M*	*SD*	*M*	*SD*	*t(22)*	*p*
Activation-Deactivation Checklist						
Clenched	2.00	0.510	2.60	0.670	–2.05	.05
Intense	2.20	0.390	2.70	0.650	–2.28	.03
Emotional Activation Scale						
Anxiety	2.50	2.03	8.60	8.36	–2.48	.02
Disgust	1.62	1.97	6.94	8.72	–2.05	.05

Note. CE = context enhanced

TABLE 2
Mean Scores, F, and P Values on the Mission Performance Subscales for the Control and Context Enhanced Groups

	Control Group		*CE Group*			
Performance Subscales	*M*	*SD*	*M*	*SD*	*F*	*p*
Mission success[a]	12.66	3.33	16.38	4.03	6.075	.022
Game adaptation[b]	24.27	3.37	26.28	5.46	0.689	.415
Use of tactics[c]	29.50	4.47	28.84	7.08	0.075	.787

[a]Subscale items: second waypoint reached; avoided taking fire; use of alternative or creative ways to accomplish mission; skilled use of time; followed through with mission and objectives

[b]Subscale items: use of provided functions of mouse and keyboard; switching between weapons; using full visual (e.g., nightvision, scope) and weapons inventory

[c]Subscale items: use of strafing (sidestepping); cautious approaches into areas; used scope and maintained distance from targets when engaging; maintaining cover while engaging

subsequently performed an additional analysis on the experimental group alone that showed no performance differences between participants who had and had not previously seen the film.

Within-Group Prediction of Performance

A stepwise multiple regression was performed on data from all participants (n = 24) to determine which, if any, indicators from the postsession ADC and EAS along with the condition assignment contributed significantly to the prediction of the "mission success" scale. For the analysis, scores on the subscale "mission success" were used as the dependant variable, and the condition assignment along with the poststress indicators from the ADC and EAS were used as independent variables. Table 3 displays the most parsimonious, highly significant model of the three models yielded. R^2 for regression was significantly different from zero, $F(3,19) = 9.07$, $p < .001$, with three variables entered predicting 59% of the variance in performance on mission success (adjusted $R^2 = .524$).

"Overall performance," the subscales taken together, was subjected to a hierarchical regression analysis (see Table 3). Of interest, 88% of the variance in "overall performance" could be accounted for by the experimental condition, along with the postvideo reports of higher energy, more tense, less clenched, less intense, and less restful, $F(6,16) = 19.01$, $p < .001$, adjusted $R^2 = .831$.

DISCUSSION

Advanced simulation-based training environments have the unique ability to produce human arousal and stress, thus offering promising contributions to the field of

training. Current simulations can provide both positive (e.g., entertainment-induced humor, energy, and/or motivation) and negative (e.g., conflict-induced fear and/or aversive physiological reactions) types of stress and arousal for a wide variety of realistic training applications. This ability is beneficial because both positive and negative stress has been shown to enhance skill retention as well as the transfer of training from the simulator to the real world (e.g., Mayer & Volanth, 1985; Williams, 1980). Advanced simulations can provide realistic conditions with stressful emergency situations and dangerously unexpected situations. Simulations, however, specifically developed to provide entertainment (as in OTS games) may inadvertently reduce the negative arousal or stress context that learners must experience during war preparation training. Specifically, simulations that are contextually defined as gaming may threaten training effectiveness in all domains that heavily rely on context-relevant stress and intensity as a formal part of the training (e.g., military, emergency responders). This study addressed the contextual conflicts of using OTS entertaining PC-based games for infantry training. By supplementing an OTS infantry game training session with an intense and vivid video depiction of a front-line infantry battle, we sought to influence the training and outcome performances of exposed participants. The success of the manipulation was confirmed because this experimental group produced significantly higher responses on self-reported stress indicators than the control group who were exposed to the unstressful video analog.

TABLE 3
Results of Regression Analyses*

Variables	*B*	β	R^2	*Adjusted* R^2	*F*	*df*	*p*
Hierarchical							
Constant	73.175		.877	.831	19.01	6,16	.000
Condition	13.626	.36					
Energetic	40.809	.994					
Tense	32.302	1.221					
Clenched	–47.962	–1.447					
Restful	–11.288	–.427					
Intense	–13.090	–.400					
Stepwise							
Constant	6.91		.589	.524	9.07	3,19	.001
Condition	5.69	.70					
Energetic	4.14	.47					
Intense	–2.72	–.39					

Note. ADC = Activation-Deactivation Checklist; EAS = Emotional Assessment Scale.

*Results of a Hierarchical Regression Model (The Most Parsimonious) Used to Predict "Overall Performance" from the ADC and EAS Stress Indicators after Entering Condition Assignment and Results of a Stepwise Regression Model (The Most Parsimonious) Used to Predict Performance on the "Success" Subscale When Considering Both the Condition Assignment and the ADC and EAS Stress Indicators

Motivation to Succeed

As expected, the experimental group had significantly higher scores on the post-training measure of "mission success." This subscale was composed of performance-related items such as completing mission objectives, number of hits, and waypoints reached. The scale represents the traditional measurements of gaming performance that are often found in the literature on practice. Conversely, the two groups did not differ significantly on the unique subscales "learned tactics" or "game adaptation." This is of special interest because the items in the "learned tactics" and "game adaptation" scales were associated with specific infantry and game training content (e.g., cover and concealment, use of equipment), whereas the "mission success" scores represented participant's performance or completion of mission objectives. Because initially there were no significant differences between the groups on items such as individual's gaming habits and initial stress, an important conclusion is that the experimental manipulation (stress context-relevant materials) produces more "motivation to succeed" in game training, but may not affect specific skill acquisition per se.

Individual Reactions to Warlike Stressors

The experimentally manipulated group produced significantly higher scores on the self-reported measures of stress overall (both positive and negative forms) as well as subsequent enhancement of mission success. However contrary to expectations, the "overall performance," or the three subscales combined, did not differ across the two groups. An explanation for this comes from similar research that suggests that if individual differences are not accounted for, performance during stressful conditions may yield ambiguous results (Driskell et al., 2001). Accounting for individual differences is important because of the differential reactions of humans to stressors. The average response observed in the "stress group" may not reflect the fact that some individuals were not effectively "stressed." To interpret the results of those participants who were effectively stressed (the manipulation was successful for) on outcome performance, a hierarchical multiple regression analysis was used. Indeed, when adding individual's stress-specific responses to the experimental manipulation, 88% of the variability in "overall performance" (the three subscales combined) was predicted. Specifically, the analysis showed that the 24 participants' overall performance was highest when they were (a) exposed to the supplementary context enhancement (the stress condition) and (b) reported higher motivational or positive stress reactions (higher energy, higher tenseness, less intensity, and less clenched).

Stress Adaptation

Overall performance (including unique measures of actual material trained) is best when exposed to a stressful condition that actually results in higher motivational or

positive stress of an individual in game-based training. The ability of an individual, however, to produce positive arousal under intense conditions varies in the population. It is suggested that individuals who can achieve and maintain an adaptation to these stressors, or are relatively positively influenced by the intensity of war like images, may be the most effective in the pursuit of rapid training and deployment in frontline infantry battle conditions. Performance outcome as a result of stress adaptation has also been supported by other studies. For example, Caterini, Delhomme, Descheumed-Moloinaro, and Dittmar (1995) described human performance outcome to be dependent on the individual's mastery of the stress (lower physiological responses) or adaptation to stimulus-induced emotions. Similarly, Brehm (1999) described adaptation as a motivational state derived by human moderation of emotional indicators. Further experimentation is required to examine the relationships between motivation from stressful conditions and stress adaptation from the conditions in an infantry context. Further research is also needed to seek reliable indices that include realistic warlike stressors and particularly measure types of stress that specific individuals experience as well as their rate of adaptation.

Summary

Advanced graphics and special effects have traditionally been suggested to increase immersion, thereby generating arousal and optimal human learning conditions, yet very little empirical evidence exists regarding the interplay of advanced realism or immersion, contextual arousal, and resultant transfer of training. To our knowledge, this is the first study to report the effects of contextually relevant supplemental material, in the form of warlike stress, to game-based training. The effect of the warlike stressor on game training and performance outcome is interpreted as an increase in performance motivation only. A change in motivation may be inferred because the scores of the experimental group on "mission success" were significantly higher than the control group, whereas the scores of actual training content (e.g., learned tactics and game use) did not differ between groups. Of interest, although increased arousal via movie-like special effects enhanced mission success, possibly through motivation, it may not have influenced specific training and retention as previously expected. At present we are also uncertain as to how long such motivating effects persist. Our experiment shows that short-term persistence is evident; however, whether this effect continues over periods of days and weeks is a phenomena that requires further experimentation.

Alternatively, additional analyses indicated that when individuals were both exposed to realistic warlike stress images and reacted with positive arousal, they effectively retained training and had higher performance scores overall. The optimal training, performance, and retention characteristics associated with response to realistic conflict should be screened for in recruits, especially when there is a need for rapid conflict training. Knowledge of these results can also aid in the design of effective adaptive instructional systems used for a wider variety of recruit charac-

teristics. The addition of context relevant stress, along with feedback from the learners as to the type of arousal occurring during training, can aid systems in adapting the context and motivational aspects to the users needs; we can thus proceed with high-stress contexts-based training at a rate fostering positive human arousal and thus effective performance.

To the present, our work has examined military naïve personnel whose transfer condition consists of SME-evaluated game mission success. With recent world events, it is evident that PC-based game training combined with effective supplementary stress might be used to assist rapid-deployment troops who will face immediate immersion in real world conditions. Although our results suggest this effect on university students, this may not be the case for military personnel. Should these findings be sustained for such personnel, the results can evidence a strong, positive impact on force readiness and response resilience. It is toward such a goal that our on-going work is directed.

ACKNOWLEDGMENT

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

REFERENCES

Baker, D., Prince, C., Shrestha, L., Oser, R., & Salas, E. (1993). Aviation computer games for crew resource management training. *International Journal of Aviation Psychology, 3*, 143–156.

Bowen, D. (1987). A theory of experiential learning. *Simulation and Games, 18*, 192–206.

Brehm, J. W. (1999). The intensity of emotion. *Personality and Social Psychology Review, 3*, 2–22.

Brown, A. H. (1999). Simulated classrooms and artificial students: The potential effects of new technologies on teacher education. *Journal of Research on Computing in Education, 32*, 307–318.

Carlson, C. R., Collins, F. L., Stewart, J. F., & Porzelius, J. (1989). The assessment of emotional reactivity: A scale development and validation study. *Journal of Psychopathology and Behavioral Assessment, 11*, 313–325.

Caterini, R., Delhomme, G., Deschaumes-Molinaro, C., & Dittmar, A. (1995). Increased activation as a limiting factor in sharp shooters. *Neurosychologia, 33*, 385–390.

Coleman, D. S., & Johnston, J. H. (1999). Applications of commercial personal computer games to support naval training requirements: Initial guidelines and recommendations. *Proceedings of the 21st Interservice/Industry Training Systems and Education Conference,* Orlando Florida, 1999.

Driskell, J. E., Johnston, J. H., & Salas, E. (2001). Does stress training generalize to novel settings? *Human Factors, 43*, 99–110.

Driskell, J. E., & Salas, E. (1991). Group decision-making under stress. *Journal of Applied Psychology, 76*, 473–478.

Flach, J. M., Hancock, P. A., Caird, J., & Vicente, K. (1995). (Eds.). *Global perspectives on the ecology of human-machine systems* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hancock, P. A. (2003). The future of simulation. In D. Vincenzi & J. Wise (Eds.), *Human factors in simulation and training*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., in press.

Hancock, P. A., & Desmond, P. A. (Eds.). (2001). *Stress, workload and fatigue*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained-attention. *Human Factors, 31*, 519–537.

Homan, W. J. (1998). Design of situational awareness training for pilots. *Educational Media International, 35*, 21–25.

Johnston, J. H., & Cannon-Bowers, J. (1996). Training for stress exposure. In J. E. Driskell & E. Salas (Eds.), *Stress and human performance* (pp. 223–256). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kantowitz, B. H. (1988). Laboratory simulation of maintenance activity. *Proceedings of the 1988 IEEE 4th Conference on Human Factors and Nuclear Power Plants* (pp. 403–409). New York: IEEE.

Mayer, J. D., & Volanth, A. J. (1985). Cognitive involvement in the mood response system. *Motivation & Emotion, 9*, 261–275.

Morris, C. S., & Tarr, R. W. (2002). Templates for selecting PC based synthetic environments for application to human performance enhancement and training. *Proceedings of the IEEE Virtual Reality 2002 Conference* (pp. 109–118). New York: IEEE.

Ricci, K., Salas, E., & Cannon-Bowers, J. A. (1996). Do computer-based games facilitate knowledge acquisition and retention? *Military Psychology, 8*, 295–307.

Rogers, C. R. (1969). *Freedom to learn.* Columbus, OH: Prentice Hall.

Tarr, R. W., Morris, C. S., & Singer, M. J. (2002). *Low cost PC gaming and simulation: Doctrinal survey* (Technical Report to the U.S. Army Research Institute, 0622785). Army Research Institute, 2002.

Thayer, R. E. (1967). Measurement of activation through self-report. *Psychological Reports, 20*, 663–668.

Williams, J. M. (1980). Generalization in the effects of a mood induction procedure. *Behaviour Research and Therapy, 18*, 565–572.

Wilson, G. P., Skelly, J., & Purvis, B. (1999). Reactions to emergency situations in actual and simulated flight. *Human Performance in Extreme Environments, 4*, 34–45.

Examination of Attentional Mechanisms Underlying Stress and Performance

J.L. Szalma, T. Oron-Gilad, & P.A. Hancock

Our program of research aims to further the development of psychological theory of stress and human performance, and to articulate potential strategies for the mitigation of stress effects in the context of combat environments. To accomplish these goals we are engaged in a series of studies to test theoretical propositions derived from the maximal adaptability model of Hancock and Warm (1989). A major contribution of this model is the explicit recognition that tasks themselves represent the proximal source of stress. Individuals cope with this stress by narrowing the focus of their attention to salient cues in the environment, including task characteristics. Hancock and Warm (1989) also specified two primary dimensions along which task parameters impact an individual's adaptation to stress. These are information structure, primarily spatial in nature, and information rate, reflecting the temporal elements of tasks. Our research has focused on testing hypotheses regarding attention to spatial and temporal task elements, and the degree to which these co-vary. Preliminary results indicate that stress does not impact these dimensions equivalently. In a related line of research, we are testing basic postulates of Fuzzy Signal Detection Theory (FSDT; Parasuraman, Masalonis, & Hancock, 2000) with the goal of applying these techniques to understanding detection performance under stress in complex real-world domains. Assessment of performance under stress has been constrained by the limitations of current performance metrics to evaluate performance in operational settings. FSDT promises to improve measurement of detection performance by capturing the uncertainty not only in the observer but also in the stimulus dimensions themselves. Such improvement will enhance our understanding of how task and environmental variables impact performance under stressful conditions. Current efforts are aimed at empirically evaluating the FSDT model and comparing it to more standard signal detection analysis techniques. Results to date indicate that FSDT is a viable tool for examination of perceptual sensitivity and response bias in target detection tasks with a high degree of stimulus uncertainty.

References

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519-537.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42*, 636-659.

PROTECTING SOLDIERS: WARNING PRESENTATION AND RETENTION UNDER STRESS IN A MILITARY TASK

Jessica Helmick-Rich, Kelly A. Burke, Tal Oron-Gilad,
Jennifer Moore, and Peter A. Hancock
University of Central Florida

The donning of protective gear is invaluable when protecting soldiers from severe injuries and even death in combat. For example, the results of a study including 118 Army troops which were evacuated from Iraq with severe battlefield injuries yielded 73 percent of the injuries were to the hands, feet arms, and legs with only 9 percent of the injuries incurred to the abdomen, chest, back, or groin (Brown, 2003). All of these soldiers were wearing body armor at the time the injuries were incurred, thus, wearing the protective gear decreased the number of severe and fatal injuries which may have occurred if the body armor was not worn. Consequently, protective gear can only protect the soldier that dons the gear. Recently, body armor has been technologically improved increasing its effectiveness when worn. Recent improvements with body armor include reducing the weight of the gear, yet for ground troops the gear may still be a fatiguing load to wear causing the soldier to remove the body armor. Although it is not always necessary for soldiers to wear all of their protective equipment, such as the protective mask, it is crucial to communicate warnings effectively for soldiers to be warned when a hazardous situation arises and what action to take to protect themselves, such as donning the appropriate protective equipment.

Another aspect that must be considered in communicating warnings effectively is that warnings are usually not presented in isolation, but are generally presented while the soldier is performing a certain operational task and often under stressful conditions. Thus, warning messages must attract the soldiers' attention, inform them of the hazard, and convince them to comply. The format in which warnings is presented is not standardized; warnings vary in their presentation format across tasks and environments. Presentations of hazard warnings are commonly found in one of the following formats: pictorial, written, or auditory. Consequently, the literature pertaining to modality is unclear as to which types of presentation is most memorable, salient, or most effective resulting in a higher rate of compliance behavior under stressful conditions.

Historically, research in the warning modality domain concentrated on comparing verbal-written presentations or picture-word differences. Penney (1975) concluded that auditory presentation (spoken words) of information results in better retention then visual presentation (written words) in short term memory tasks. Others suggested that people can recognize and recall pictures more quickly and accurately than words (Ells & Dewar, 1979; Janda & Volk 1934; Paivo, Rogers, & Smythe, 1968; Standing, Conezio, & Haber, 1970). Several theories support the claim that pictures are processed both verbally and symbolically, therefore having an advantage over words (Pavio, 1971; 1975; & 1978; Siegel & Dhawan, 1975). Similarly, literature in text processing found pictorials to have a processing advantage over text, thus arguing that pictures facilitate comprehension (Glenberg & Langston, 1992; Larkin & Simon, 1987). Some argue that pictorials are "instant reminders" of the hazard (Peters, 1984), more recognizable than words, (Paivio, Rogers, & Smythe, 1968; Standing, Conezio, & Haber, 1970), and they convey information more rapidly and effectively than verbal messages (Dorris &

Perswell, 1978). Additionally, OSHA and NIOSH guidelines suggest adding a pictorial to the warning label. Contrary to the picture superiority findings, others (Robinett & Hughes, 1984; Smith & Magee, 1980) found that icons are sometimes complex and cannot convey the appropriate information without a written message accompanying them.

Clearly, modality of presentation an important factor leading to compliance behavior, however, compliance behavior may also be affected by stress. According to the Hancock and Warm model (1989), performance will be affected by stress when it increases to the point that it is outside of the comfort zone. Under low stress conditions participants are more likely to comply than under high stress conditions (Magurno & Wogalter, 1994). In the current study, operational stress level was manipulated by the task demands of the warning-compliance task (the number of association cues that the participant had to retrieve). In particular, it was of interest to determine the amount of task demand which still provides a high level of compliance. Therefore, it was hypothesized that the comfort zone would be maintained when a participant was presented with a maximum of four stimuli in any modality. Furthermore, when participants were presented with more than four stimuli they would fall outside of the comfort zone resulting in degraded performance.

In the current study we manipulated warning presentation and warning retention. Presentation was manipulated by the modality of presentation of the warning cues (verbal, written or pictorial). Retention was manipulated by creating associations between the hazard and the warning cue through different modalities while simultaneously interacting with a simulated military task. Hence this study covers not only the effect of modality on warning compliance but also the effect of working memory associations (spatial and verbal associations) on warning retention in a dual task paradigm. Our current study examines the effects of cross modality of warning presentation and retention in a dual task paradigm in a simulated military environment. In general, attentional resources are utilized better when divided across modalities (auditory and visual stimuli, for example) rather than displayed via two auditory or two visual channels, (Wickens, 1992; Wickens, Sandry, & Vidulich, 1983). Similarly, several models of working memory are associated with different modalities or different types of representations and are consistent with the notion of multiple resource pools (Daneman & Tardif, 1987; Martin, 1993; Monsell, 1984; Shah & Miyake, 1996). Since the operational task in this study is predominantly a visual and spatial task, one can hypothesize that participants would have a significantly higher rate of compliance behavior on the warning-compliance task when the warning was presented in verbal compared to written and pictorial modality because the warning-compliance task would have less interference with the operational task.

PROTECTING SOLDIERS: WARNING PRESENTATION AND RETENTION UNDER STRESS IN A VIRTUAL MILITARY TASK

Jessica Helmick-Rich, Kelly A. Burke, Tal Oron-Gilad, Jennifer Moore, & Peter A. Hancock
University of Central Florida

INTRODUCTION

The donning of protective gear is invaluable when protecting soldiers from severe injuries and even death in combat. For example, the results of a study including 118 Army troops which were evacuated from Iraq with severe battlefield injuries yielded 73 percent of the injuries were to the hands, feet arms, and legs with only 9 percent of the injuries incurred to the abdomen, chest, back, or groin (Brown, 2003). All of these soldiers were wearing body armor at the time the injuries were incurred, thus, wearing the protective gear decreased the number of severe and fatal injuries which may have occurred if the body armor was not worn. Although it is not always necessary for soldiers to wear all of their protective equipment, such as the protective mask, it is crucial to communicate warnings effectively for soldiers to be warned when a hazardous situation arises and what action to take to protect themselves, such as donning the appropriate protective equipment.

Another aspect that must be considered in communicating warnings effectively is that warnings are usually not presented in isolation, but are generally presented while the soldier is performing a certain operational task and often under stressful conditions. Thus, warning messages must attract the soldiers' attention, inform them of the hazard, and convince them to comply. The format in which warnings is presented is not standardized; warnings vary in their presentation format across tasks and environments. Presentations of hazard warnings are commonly found in one of the following formats: pictorial, written, or auditory. Consequently, the literature pertaining to format is unclear as to which types of presentation is most memorable, salient, or most effective resulting in a higher rate of compliance behavior under stressful conditions (Ells & Dewar, 1979; Glenberg & Langston, 1992; Janda & Volk 1934; Larkin & Simon, 1987; Paivo, Rogers, & Smythe, 1968; Penney, 1975; Standing, Conezio, & Haber, 1970).

In the current study, we manipulated warning presentation and warning retention in a dual task paradigm. Presentation was manipulated by the format of presentation of the warning cues (verbal, written or pictorial). In order to determine which type of presentation is most effective, we measured retention performance by calculating which format had the highest recall rate. Retention performance of the learned associations between the hazard and the warning cue was measured by the number of correctly recalled associations while simultaneously interacting with the simulated military task. Furthermore, in accordance with the Wickens' multiple resource theory (Wickens, 1984; Wickens, Sandry, & Vidulich, 1983), it was hypothesized that warnings presented in verbal compared to written and pictorial formats would result in significantly higher rate of compliance behavior since the simulated military task in this study is predominantly a visual and spatial task.

Clearly, format of presentation is an important factor leading to compliance behavior; however, compliance behavior may also be affected by stress. An experiment conducted by Magurno and Wogalter (1994) concluded that participants under low stress conditions were more likely to comply with warnings than under high stress conditions. Thus, these results are in accordance with the Hancock and Warm model (1989). This model predicts that performance will be affected by stress when it increases to the point that it is outside of the comfort zone. In the current study, operational stress level was manipulated by the task demands of the warning-compliance task (the number of association cues that the participant had to retrieve). In particular, it was of interest to determine the amount of task demand which still provides a high level of compliance. Therefore, it was hypothesized that the comfort zone would be maintained when a participant was presented with a maximum of four stimuli in any format. Furthermore, when participants were presented with more than four stimuli they would fall outside of the comfort zone resulting in degraded performance.

Compliance to warning information may also be affected by individual differences in working memory capacity. If an individual does not have the resources to support the amount or type of warning information that is to be stored or processed, then they may not be capable of heeding the warning. Many theorists support the notion of separability of working memory resources for spatial thinking and language comprehension (Daneman & Carpenter, 1980; Daneman & Tardif, 1987; Shah and Miyake, 1996). In accordance with the separability of working memory theory, this experiment seeks to determine whether warning presentation is contingent upon individual differences in available capacity and processing resources in verbal and spatial working memory. It is hypothesized that the processing and storage capacity will determine an individual's ability to comply with warnings presented in the different modalities.

METHOD

Participants

Twenty-two (mean age = 21.2, *SD* = 3.4*)* participants were recruited on a voluntary basis from undergraduate psychology classes at the University of Central Florida and were paid approximately $37.50 for their participation (based on an hourly rate of $7.50).

Materials

Dual Task Setting. The Warning Color-Combination (WCCOM) compliance task and the shooting task were presented on two separate monitors with two keyboards and mice (See Figure 1).



Figure 1. Experimental set-up.

WCCOM Compliance Task. WCCOM was one of the two tasks in the dual task paradigm. The warning-color combinations in this task consisted of one of ten warnings (boots, earmuffs, glasses, gloves, helmet, shield, suit, respirator, meter, or mask) paired with one of ten colors (red, blue, green, orange, purple, black, white, gray, brown, or yellow). The WCCOM consisted of both a storage and a processing requirement. The storage requirement consisted of learning the color associated with each warning. Warnings were presented in one of three modalities, pictorial, written, or verbal. Each combination was paired randomly and appeared only once per block.

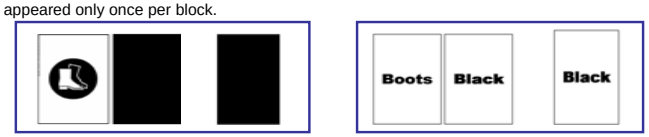


Figure 2. Example of a pictorial warning-color combination (left) and the color stimulus (right) that elicits the key press response during the WCCOM portion of the dual task. In this specific example, the warning, boots, are combined with the color black (left). Example of a written warning-color combination (left) and the color stimulus (right) that elicits the key press response during the WCCOM portion of the dual task (right).

Each warning-color combination was presented individually on the computer screen for five seconds. An example of the pictorial presentation of the WCCOM is depicted in Figure 1. In this example the warning, boots, is paired with the color black. The written warnings were presented in the same manner. The verbal WCCOM was presented via speakers. For example, the participants heard "*boots...black*". Operational stress level was manipulated by the working memory demands of the WCCOM, the number of association cues that the operator had to retrieve, which increased in demand from two to four to eight.

Shooting Task. During each trial, the participants completed a two-minute mission in Ghost Recon. During each mission, the objective was to clear a building by entering each room and killing all of the enemies. Performance was measured by calculating the hit percentage (number of shots which hit a target divided by the total number of shots fired).

Procedure

- An all within participants design
- Three blocks (one for each modality: verbal, written, and pictorial); 15 trials
- Presentation order of the modality conditions were counterbalanced and each block was randomized.
- Testing occurred in two sessions (approximately 2.5 hours each)
- Session 1
  - informed consent, demographic questionnaire, four working memory tasks
  - practice session
  - first block.
- Session 2
  - Second and third block of
  - During both sessions, the Rating Scale Mental Effort & NASA-Task Load Index

RESULTS

WCCOM Task. A three-way 3 (modality) X 3 (task demand) X 5 (trial) X 2 (response type) repeated measure analysis of variance (ANOVA) was conducted on warning compliance behavior. The within subjects variables were modality, task demand, and trial. The between subject variable was response type.

- There was a significant main effect of modality and task demand, $F(2, 19) = 21.9$, $p < .0005$, partial $\eta 2 = .70$, $F(2, 19) = 9.4$, $p < .0005$, partial $\eta 2 = .50$, respectively. No main effects for response type or trial were found.
- Post hoc tests for modality type show that participants were significantly more likely to comply when the information was presented in the written, ($M = 67$, $SD = 3$) than in pictorial ($M = 56$, $SD = 3$) or verbal modality ($M = 54$, $SD = 3$).
- Three two-way interactions were found. The first interaction was between modality and response type $F(2,19) = 11.4$, $p < .0005$, the second between task demand and response type $F(2,19) = 63.3$, $p < .0005$, and the third between task demand and trial $F(8,13) = 1.03$, $p < .0005$.

Workload. A three-way 3 (modality) X 3 (task demand) X 5 (trial) X 2 (response type) repeated measure analysis of variance (ANOVA) was conducted on RSME workload scores. The within subjects variables were modality, task demand, and trial. The between subject variable was response type.

- There was a significant main effect of modality and task demand, $F(2, 19) = 6.98$, $p = .005$, partial $\eta 2 = .42$, and $F(2, 19) = 16.9$, $p < .0005$, partial $\eta 2 = .64$, respectively. No main effects for response type or trial were found.
- Post hoc tests for working memory show that participants were significantly more likely to comply at level two ($M = 55$, $SD = 3$) than at level four ($M = 55$, $SD = 3$) and eight ($M = 57$, $SD = 3$), but not between levels of four and eight.
- Two two-way interactions were found. The first interaction was between task demand and response type $F(4,17) = 6.55$, $p = .007$, and between response and trial $F(4,17) = 6.27$, $p = .003$.

NASA-TLX scores are currently being analyzed.

Ghost Recon Shooting Task. A three-way 3 (modality) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on percentage of enemies killed compared to the number of enemies in that mission.

- There were no significant main effects or interactions.

Individual Differences in Working Memory Tasks. Data collection is ongoing; a larger sample size is needed in order to analyze the data**.**

DISCUSION

The results of this study are relevant in securing the safety of soldiers. This study identified the best format and task demand to present warning effectively to increase the likelihood of soldiers wearing protective gear. Results of this experiment indicate that participants presented with warnings in written or pictorial format complied significantly more than when presented with verbal warnings regardless of response mode. In addition this mode of presentation generated the least workload on operators. Consistent with the Hancock and Warm model of stress and attention (Hancock & Warm, 1989), comfort zone was maintained when a maximum of four stimuli in any format of presentation. Furthermore, when eight stimuli were presented, retention degraded. Thus, when task demands (task stress) are relatively low (two and four warning stimuli), the format of presentation is of less importance and participants are able to comply. When task demands are relatively high (eight warnings) the format of presentation was important and played a significant role in compliance behavior.

The implication that task demand has on the modern day soldier is critical to their survival. With the increase of information presented to soldiers, they are often presented with critical warning information while simultaneously performing some other task or receiving other important information. The warning message must be salient, and stand out from other environmental distractions. Thus, when task demand is high, the presentation format of warning information needs to be considered.

Future research in this area will first determine whether warning presentation is contingent upon individual differences in working memory storage and processing capacity. In this current line of research a continuation of collecting data on individual differences will continue. Secondly, the current study only investigated the differences between pictorial and written response modes on the WCCOM task, in the future a verbal response mode will be implemented and compared.

**This research was facilitated by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ARO under grant DAAD19-01-1-0621. This work was also supported in part by DARPA grant NBCH1030012, Program Manager, Dylan Schmorrow.*

24th Army Science Conference
Transformational Science and Technology for the Current and Future Force

Mapping an Objective State of the World to a Subjective Degree of Signalness: an exercise in Fuzzy Signal Detection Theory Analysis

Jared R. Kirkland, Tal Oron-Gilad & Peter A. Hancock
University of Central Florida

This paper addresses the improvements that may be made upon classic signal detection theory (SDT) by fuzzy SDT. A stimulus set is constructed upon which a fuzzy analysis may be run. The first step in a fuzzy analysis of data, constructing a mapping function of the objective state of the world (SOW) onto a subjective function, is detailed.

Introduction

Classic signal detection theory (SDT) is one of Behavioral Science's most robust tools for measuring human performance (Wickens, 2002). Since its popularization in the 1960s by Swets and Tanner (Swets, 1964; Green and Swets, 1966), SDT has been verified beyond question as to its value to theoretical and practical psychophysics and human performance measurement. However, three researchers have recently pointed out some weaknesses inherent in this tool and have offered resolutions to strengthen it (Parasuraman, Masalonis, & Hancock, 2000; Hancock, Masalonis, & Parasuraman, 2000). These improvements lie in the ability to capture signal and response variability across contexts and over time in a detection performance analysis.

Fuzzy signal detection theory (FSDT) (Parasuraman, Masalonis, & Hancock, 2000; Hancock, Masalonis, & Parasuraman, 2000) is the marriage of fuzzy set theory (Zadeh, 1965) and SDT. Fuzzy set theory recognizes that the presence of a given signal may range anywhere between absolute presence and absolute absence, combined with the variable presence of noise. FSDT is sensitive to the obscurity within which many real-world signals are embedded, and allows for responses to range on a similar continuum from one hundred percent certainty of signal presence to an equal level of certainty regarding the absence of a signal. Incorporating all of the possible levels of signalness and response certainty into a detection model yields a more accurate snapshot of what is occurring in any given signal detection trial. Furthermore, capturing this variability of signal and response certainty not only at a given moment, but also across contexts and over time yields a truer reflection of real-time human performance as a function of the state of the world (SOW). This is what FSDT promises to accomplish.

The theory behind FSDT is solid and its prospects for improving human performance (be it detection, decision making, etc.) measurement are exciting. However, because it is new, there have been only a relative handful of studies run to validate this tool. The goal of the current study, therefore, is to create a stimulus set against which the basic postulates and theories of FSDT can be validated. Steps to this end include developing a psychophysical function for a set of morphing videos and mapping the objective state of the world, or in terms of this study the amount of signal present in a given video frame, to the subjective degree of signalness (see Table 1). This is the first of four steps in running a fuzzy analysis. For details regarding this and the remaining three steps, I

refer the reader to Parasuraman, Masalonis, & Hancock (2000) and Hancock, Masalonis, & Parasuraman (2000), the founders of this theory.

Method

The stimuli for this study were generated using Morph Man v.4, a morphing software developed by STOIK Imaging, and presented in video format. Two tanks, an American M1A1 Abrams and an Iraqi T55, were used as morphing objects. Morph Man morphs the initially presented tank, say an M1A1, into a T55 by gradually replacing M1A1 characteristics with T55 features. This is done in a global manner such that all features of the tank are morphed simultaneously. To further explain this process, the object that a participant initially views is a tank with 100% M1A1 features. A quarter into the video, the object contains 75% M1A1 features and 25% T55 features. At the halfway point of the video, the object contains an equal portion of features from both tanks. This process continues in a linear fashion until the end object is a tank with 100% T55 features. (For clarity on this process, Figure 1 presents a series of pictures beginning with the M1A1 followed by morphed images at 20% intervals. The final picture is a 100% T55 tank.) This stimulus set captures the obscurity that exists in real-world signals, which obscurity is a pillar in the theory behind FSDT. At all points in the video, except for the first and last frames, the object viewed by participants contains a mixture of noise, the initial tank features, and signal, the terminal tank features.



Figure 1 A morphed tank sequence beginning with an M1A1. These pictures progress in 20% increments and end with a T55.

The task involves participants watching videos as detailed above. These morphing videos present tanks at two positions, a left side view and a right side view. According to traditional method of limits procedures (Myers & Well, 2003), participants will be instructed to indicate when the initially presented tank begins to look like the terminal tank. In terms of signal detection, participants will be asked to respond when enough of

the signal, or features of the terminal tank, is present to conclude that the current image contains a greater portion of terminal tank features than initial tank features. As soon as the participant presses the spacebar to indicate their perception of a change from initial to terminal tank, a second video will run beginning with the terminal tank morphing into the initial tank. This cycle will be repeated nine times, for a total of ten trials, before moving on to the second tank angle.

To guard against errors of habituation and expectancy (Myers & Well, 2003), three video lengths, 6.6 seconds, 10.0 seconds, and 13.3 seconds, are used. The order of presentation is dictated by a Latin square, and initial tank type is counterbalanced among participants.

The experiment is being run with Inquisit by Millisecond Software. At each response, Inquisit records the elapsed time from stimulus onset. This elapsed time is then used to determine which frame contained enough of the signal for the participant to respond. This data is then used to create a mapping function as detailed in the Discussion section.

Twenty college undergraduates attending the University of Central Florida will be paid $7.50 per hour to participate in this study. It is estimated that each participant will require one hour to complete the study.

Results

Based on pilot data, we expect that participants will not detect a shift in tank characteristics at the video's halfway point. Further, we expect that participants will respond 70 to 75 percent into the video, or when 70 to 75 percent of the terminal tank features, the signal, have replaced the initial tank characteristics. This outcome will occur regardless of which tank is presented as the initial tank.

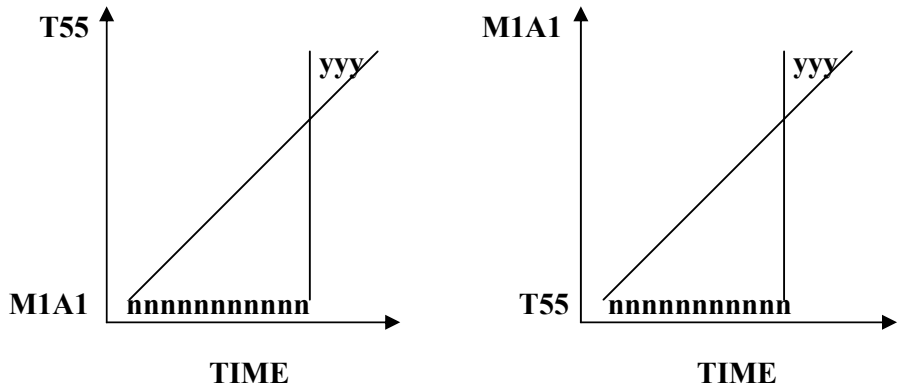


Figure 2 An hypothesized function representing when participant recognize a change from initial to terminal tank.

Discussion

Again, the purpose of this study is to understand the psychophysical function of this particular stimulus set. Once we obtain this understanding, we will be able to map the objective SOW onto a subjective continuum of perception. (An example of how this might turn out is presented in Table 1.) For example, suppose that the obtained mean for when the initial tank begins to look like the terminal tank falls at the objective 65 percent mark. In the follow-up study to this work, which will be a full FSDT study, the objective 65 percent image will be treated as containing equal portions of both tanks. This process will be carried out so that the entire range of objective signalness is mapped onto the full range of subjective signalness from 0.00 to 1.00. We will then be able to use single frames from the morphing process in an FSDT task with the end goal of validating this new tool for measuring detection performance in particular and human performance in general.

TABLE 1: An hypothesized mapping function from the morphing tanks stimuli

Objective degree of signalness	Perceived deg. of signalness	Objective degree of signalness	Perceived deg. of signalness
0.00	0.00	0.55	0.10
0.05	0.00	0.60	0.25
0.10	0.00	0.65	0.50
0.15	0.00	0.70	0.65
0.20	0.00	0.75	0.75
0.25	0.00	0.80	0.80
0.30	0.00	0.85	0.85
0.35	0.00	0.90	0.90
0.40	0.00	0.95	0.95
0.45	0.02	1.00	1.00
0.50	0.05		

Note: Increments of .05 under objective degree of signalness was used for space considerations.

References

GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

HANCOCK, P. A., MASALONIS, A. J., PARASURAMAN, R. On the theory of fuzzy signal detection: theoretical and practical considerations. *Theoretical issues in Ergonomics Science*, **1**, (3) 207-230.

MYERS, J. L., and WELL, A. D. (2003). *Research design and statistical analysis.* Mahwah, New Jersey: Lawrence Erlbaum.

PARASURAMAN, R., MASALONIS, A. J., and HANCOCK, P. A. (2000). Fuzzy signal detection: basic postulates and formulas for analyzing human and machine performance. *Human Factors*, **42**, (4) 636-659.

WICKENS, T. D. 2002, *Elementary Signal Detection Theory* (New York: Oxford University Press, Inc.).

ZADEH, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office

(Grant# DAAD19-01-1-0621). The views expressed in this paper are those of the author and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the U. S. Government.

Are There Differences in Teamwork Across Team Type?

Heather A. Priest, Joseph Guthrie, C. Shawn Burke, Clint Bowers*, & Eduardo Salas
Department of Psychology, Institute for Simulation and Training, * Team Performance Lab
University of Central Florida
Orlando, FL

Team performance is a complex and illusive phenomena that has not yet been fully understood. Teams rely on certain underlying processes to shape their performance. These underlying processes (i.e., "the black box") have been investigated over the last few years. However, a significant amount of the research has relied on command-and-control teams. The current study focuses on the differentiation of teams based on interdependence, performance arrangement and shared awareness to test the relationship between performance and teamwork behaviors. In addition, the effects of stress, operationalized as workload using a secondary task was also tested. Results showed some support for different communication patterns in different team types. No effects of workload were found. These results should be further investigated and applied to future team research.

The focus of traditional team research has revolved around performance and the measurable tangibles that teams deliver. However, within recent years there has been a move to understand the 'black box' known as team process. Beyond the observable outputs of teams (e.g., products), teamwork consists of unseen processes (for more information on team processes see Salas, Burke, & Cannon-Bowers, 2000). These processes are the foundation for effective teamwork and must be understood in order to promote favorable team outcomes (Driskell & Salas, 1992). This line of research alludes to a longstanding question in psychology: how do you measure the unobservable? One answer to this with regards to teams is communication. Recent literature (e.g., Bowers, Jentsch, Salas, & Braun, 1998; Orasanu, 1990; Thornton, 1992) has attempted to measure teamwork and the underlying processes by capturing and analyzing team communication.

One method of communication analysis (and the application used within the current study) is Team Dimensional Training (TDT; Smith-Jentsch et al, 1996). TDT systematically identifies teamwork behaviors that are essential for good team performance. TDT is structured around four teamwork dimensions: 1) information exchange, which involves knowing what information to pass to team members and when to pass the information; 2) communication delivery, which focuses on how the information is delivered; 3) backup behavior, which involves team members taking action to help alleviate some of the workload on other members of the team; and 4) initiative/leadership, which involves behaviors that help to guide or provide direction for the team. These four dimensions were developed by comparing the behaviors of experienced teams with those of inexperienced Navy teams. Each of the four above dimensions contains a subset of specific teamwork behaviors (Table 1).

While methods for examining the 'black box' (e.g., communication analysis, mental model measurement techniques) have shown just how far team research has come, there are still a number of issues to be explored. One of these and the focus of the current study is the question of how different team types may vary with respect to application of teamwork processes. For example, does a successful command -and-control team use the same processes as a successful combat team? Do distributed teams function in the same way as co-located teams? The essence of these questions addresses a more practical issue facing those of us that study and use teams: can the findings of the last three decades be applied to all kinds of teams?

Information Exchange	Communication Delivery
• Utilizing all available sources of information • Passing information to the right persons without having to be asked • Providing 'big picture' updates	• Proper phraseology • Completeness of standard reports • Brevity/avoiding excess chatter • Clarity/avoiding inaudible communications
Supporting Behavior	**Initiative/Leadership**
• Monitoring & correcting errors • Providing & requesting backup or assistance to balance	• Providing guidance or suggestions • Stating priorities

Table 1: Categorization of TDT Teamwork Behaviors (Smith-Jenstch et al., 1998)

Team research has provided organizations and the military with a plethora of "how to's" for successful teams through the years (e.g., Salas, Cannon-Bowers, & Johnston, 1997). However, the majority of research relies on command-and-control teams; therein the danger lies in the limit of psychologists to test theories across team types. While there has been a great deal of research into different team taxonomies (e.g., Sundstrom, 1999), these typologies have focused on the taskwork of teams and have not delved into the processes these teams use to accomplish their tasks. Similar to suggestions in Cohen and Bailey (1997), we believe that with all the advances of the science of teams, it is time to examine the assumption that a team is a team is a team with regards to process.

Indicative of this trend is the fact that the most used testbeds in team research are DDD (Dynamic Distributed Decision-Making, Kleinman, Pattipati, Luh, & Serfaty, 1989), TIDE² (Team Interactive Decision Exercise for distributed

expertise, Hollenbeck, Sego, Ilgen, Major, Hedlund, & Phillips, 1997), and TANDEM (Tactical Naval Decision Making System; Kozlowski, 1996), all of which are representative of command-and-control teams. These results are typically taken and applied to any collective classified as a team according to the broad definition accepted by today's team researchers: two or more individuals interacting interdependently and adaptively towards a common goal (Salas, Burke, & Cannon-Bowers, 2000).

In an effort to begin to investigate how processes (e.g., teamwork KSA's) might differentially relate to teamwork behaviors and the relationship between behaviors and team performance across teams with different characteristics, we created a testbed that would enable investigation of this question. Specifically, following the rationale depicted in taxonomic literature on team type it was our belief that teams having different performance arrangements, levels of shared visualization, and interdependencies would see differential patterns of teamwork behavior as a function of their verbal communications in relation to their team performance. Therefore, we created a testbed where 3 team types could be represented with respect to these variables.

Another focus of the current investigation was to determine how stress may affect teams based on team type. Recent military and organizational contexts are characterized by complexity, ambiguity, and other stressors (i.e., workload, time pressure). While much is known about the impact of stress on individuals (Griffith, 1997; Hancock and Warm, 1989; Milgram Orenstein, & Zafrir, 1989; Smit, & Schabracq, 1997), there is much less known about how stress impacts teams (i.e., team process, team cognition, team performance, Cannon-Bowers & Salas, 1990; Driskell & Salas, 1992; Fiedler, Potter, Zais, & Knowlton, 1979). As teams have become a dominant force with many organizations (e.g., public, private, and government/military) over the past thirty years there is a need to address the problem of teams and stress. This is especially problematic for organizations that operate in mission critical high impact environments such as the military. More specifically, technological advances and changes in the geopolitical arena are causing the impact that stress has on both individuals and teams to become of paramount concern. Therefore, a secondary focus of this study was to investigate the effects of stress (i.e., workload) across team type.

Based on this, we developed three hypotheses addressing how teams might apply teamwork behaviors and the impact that team type and workload may have on these relationships. First, we predict that workload will have a negative effect on all teams, regardless of team type. We base this prediction on previously published literature on stress effects (Weaver, Bowers, & Salas, 2001). Furthermore, we predict that task type will moderate the relationship between stress and teamwork behaviors. For example, teams that are co-located have team level interdependence, and high shared visualization may exhibit less information exchange under high workload as their team characteristics facilitate the use implicit communication. (e.g., Ghost Recon, see method). This is in contrast to teams with distributed expertise and location with moderate levels of shared visualization where explicit communication is needed as members may not have as much common ground.

Second, we predict that teamwork behaviors at a global level may be differentially related to performance dependent on team type. While the majority of team research tends to collapse teamwork (as operationalized by team communication) into one overarching category (e.g., Mathieu, Heffner, Goodwin, & Salas, 2000; Minionis, 1994; Travillian, Volpe, Cannon-Bowers, & Salas, 1993; Zaccaro, Parker, Marks, Burke, Higgins, & Perez, 1997) we seek to push our knowledge further by examining how four broad dimensions that have been shown to discriminate (i.e., information exchange, communication delivery, supporting behavior, initiative/leadership) may vary according to team type. For example, distributed teams with sequential interdependency and low shared visualization (e.g., Flashpoint, see method section) will show a significant relationship between explicit information exchange and performance as the low level of shared visualization and distributed nature makes it essential that the team exchange information verbally to perform well. In contrast, in teams with high levels of shared visualization whom are co-located and highly interdependent information exchange may not be the best predictor of performance as teams can more easily rely on implicit communication. As another example, supportive behavior (i.e., backup behavior) will be related to performance in teams with team level interdependence that are co-located (e.g., Ghost Recon), but not in teams that are distributed with sequential or reciprocal interdependency (e.g., Flashpoint, DDD, see method). We are not, however, saying that supportive behavior is not important in all teams. Our argument is based on the definition of TDT backup behavior, which is limited to physical supportive behavior. Within a distributive environment it is very difficult to provide this type of behavior.

Third, we predict that it may be beneficial to examine teamwork behaviors at a more micro level. Specifically, we predict that the way in which broader teamwork dimensions are enacted will differentially be related to performance according to team type. For example, For example, information exchange may not be significantly related to co-located team performance at a broad level since they have more of an opportunity to communicate implicitly. However, we may find that successful co-located teams do explicitly provide 'big picture' updates to fellow team members.

Method

Participants were 96 undergraduate students (48 two-person teams) at a large southeastern university that were randomly assigned to teams and conditions. Researchers used a 3 (team type) x 2 (workload) between subjects design where participants performed as members of one of three unique team types (command-and-control, search and rescue, combat) under one of two workload conditions.

Based on the literature discussed earlier, researchers used 3 testbeds that would elicit 3 'team types' based on level of interdependency (Saavedra et al., 1993), members being co-located/distributed, and degree of shared visualization offered through the environment. Specifically, DDD was used to rep-

resent a command-and-control team (reciprocal interdependence, distributed expertise, moderate shared awareness) and modified off-the-shelf computer games (COTS) were utilized to create testbeds which represented a combat team (Ghost Recon: team interdependence, co-located, high shared visualization) and a search and rescue team (Flashpoint: sequential interdependence, distributed, low shared visualization), respectively. Mission, task, and team characteristics within each testbed were adapted so they would be representative of the three above mentioned team types.

The second between subjects variable, workload, was manipulated by creating two levels (low, high) via the use of an auditory dual task configuration. Teams placed in the high workload condition were asked to respond to requests for current position at regular intervals throughout the mission. These responses were recorded and used for analysis.

Outcome variables were broadly classified as performance. For DDD, the number of targets identified by the team was indicative of performance. In Flashpoint, participants were assessed on the number of items on a predefined list that they could confirm correctly as a team. Performance in Ghost Recon was operationalized as coordinated strategy in that team members were assessed on how they coordinated fire effectively to eliminate the enemy and complete the objective while not wasting ammunition, based on their weapon assignments.

After completing informed consent and background questionnaires, participants received task related training. Following the training, participants completed a 30 minute mission with specific task objectives based on team type. The completion of the task objectives was scored to establish comparable performance measures across testbeds, as outlined above. In addition, communications between team members was recorded during performance for coding and analysis (using TDT; Smith-Jentsch et al., 1996). Three graduate students were trained in the use of the coding scheme. Once inter-rater reliability was reach the students, blind to condition, coded tapes into TDT categories.

Results

First, we wanted to determine the impact that workload had on teamwork and corresponding task performance, according to Hypothesis 1. Upon examination of the recorded tapes the researchers found that teams placed in the high workload condition generally did not respond to the secondary task asking for current position. Instead offloading their secondary task and focusing on their primary task. Therefore, future analyses did not include workload. Potential explanations for why teams ignored the secondary task are given in the discussion/conclusion section.

To test Hypotheses 2, researchers used regression analysis to predict performance from the 4-major teamwork categories (e.g. information exchange, communication deliver, backup behavior and initiative/leadership; Table 2).

From these results, we found partial support for hypothesis 2 in that all three types of teams had performance partially predicted by information exchange. The beta weights in Table 3 show that while Flashpoint team performance was significantly related to information exchange, t(15) = 2.634, p= 0.02, both the Ghost Recon , t(15) = -0.513, p= 0.616, and the DDD team performance, t(15) = 1.514, p= 0.152, were not. There was also an overall lack of communication delivery being coded in the teams' teamwork behaviors across the three testbeds.

	Variable	B	SE B	β
Ghost Recon				
	Information Exchange	-0.018	0.035	-0.136
	Supportive Behavior	-0.406	0.267	-0.376
	Leadership Initiative	-0.051	0.087	-0.156
Flashpoint				
	Information Exchange	0.046	0.017	0.576*
	Supportive Behavior	0.341	0.567	0.159
	Leadership Initiative	0.090	0.047	0.457
DDD				
	Information Exchange	0.007	0.005	0.375
	Supportive Behavior	-0.024	0.394	-0.016
	Leadership Initiative	-0.093	0.055	-0.409

Table 2: Regression Analysis for TDT Dimensions *Note*: Communication not analyzed due to lack of data points.

* p<.05

In addition, supportive behavior was not a significant factor in Ghost Recon performance. The beta weights in Table 3 show that Ghost Recon team performance was not significantly related to supportive behavior, t(15) = -1.519, p= 0.151. Neither ghost Recon nor Flashpoint teams' performance was related to leadership/initiative, t(15) = -0.590, p= 0.565 and t(15) = 1.923, p = 0.075 respectively. Results also showed that DDD team performance was not significantly related to leadership/initiative, t(15) = -1.675, p= 0.116.

To test the third hypothesis regarding micro-level teamwork behaviors, we first evaluated the frequency with which each team type used the separate subdimensions. We used analysis of variance to investigate whether team type differentially impacts the types of teamwork behaviors used. Table 3 shows the mean frequency for each of the coded behaviors by testbed. The analysis showed that frequency of teamwork behavior was significantly different between testbeds on several of the TDT. For example, the information exchange subdimensions of seeking info from all sources, F(2,45) = 53.321, p<0.001, passing info before being asked, F(2,45) = 17.18, p<.001, and responding to requests for info, F(2,45) = 20.589, p<.001, all varied between testbeds.

The next step of data analysis was to use regression analysis using the TDT sub-dimensions as predictors of performance. Specifically, we wanted to investigate the possibility that certain sub-dimensions of the TDT teamwork behaviors were related to performance even though the major classifications were generally found not to be related to performance (Table 4). For example, while information exchange was not

significantly related to performance in neither Ghost Recon, sub-dimensions of information exchange were, with seeking information and responding to requests effecting performance in Ghost Recon, $t(15) = 3.907$, $p = 0.002$ and $t(15) = 3.223$, $p = 0.007$.

		GR	FP	DDD
Information Exchange	Seek info from all sources	76.3	91.12	20.43**
	Pass info before asked	143**	187.62**	81.37**
	Respond to request for info	57.5**	86.75**	18.68**
	Providing 'big picture' updates	0	0	.0625
Comm Delivery	Avoid excess chatter	0	.1250	0
	Ensure comms are audible/clear	0	.0625	0
Supportive Behavior	Correcting team errors	1.1875	.3125	.3125
	Providing/requesting back-up when needed	7.56*	.93	.43
Leadership Initiative	Providing guidance or suggestions	49.75	59.81	2.62**
	Stating clear team and individual priorities	.1875	0	2.12

Table 3: Differences in Mean Frequency of TDT Behaviors by Testbed *Note:* Sub-dimensions with no occurrence in any of the testbeds were omitted from the analysis. *p<0.05 **p<.001

Discussion

The purpose of this study was to determine if different types of teams exhibit different team processes in order to complete team goals and to determine if these processes change with teams under stress. In order to accomplish this, the researchers developed testbeds that were representative of 3 types of teams (command-and-control, search and rescue and combat). Participants were randomly assigned to one of two workload conditions within each testbed. The secondary task used to measure workload in the study was predominantly ignored by the participants. It is possible that the primary task exacted a large enough toll on the mental resources of the team that the secondary task was ignored in order to prevent mental overload and a decrement in performance.

The results of the ANOVA on the frequency of teamwork behaviors across the 3 types of teams suggests that indeed different types of teams employ different types of teamwork behaviors to accomplish their tasks based on the level of interdependency, level of shared awareness, type of task and roles of the team members. For example, teams in both Ghost Recon and Flashpoint were more likely to provide guidance or suggestions to their teammates than teams participating in the DDD task. This is most likely a reflection of the level of interdependency found in the teams.

	Variable	B	SE B	β
Ghost Recon				
Information Exchange	Seeking info from all sources	-0.901	0.23	-2.37*
	Passing info before asked	0.055	0.035	0.288
	Responding to request for info	0.976	0.303	1.943*
Supportive Behavior	Correcting team errors	-1.838	1.387	-0.330
	Providing or requesting backup when needed	-0.393	0.267	-0.367
Leadership Initiative	Providing guidance or suggestions	-0.049	0.091	-0.147
Flashpoint				
Information Exchange	Seeking info from all sources	0.141	0.077	0.383
	Passing info before asked	0.129	0.052	0.571*
	Responding to request for info	-0.005	0.028	-0.046
Supportive Behavior	Correcting team errors	0.05	1.426	0.01
	Providing or requesting backup when needed	0.397	0.639	0.17
Leadership Initiative	Providing guidance or suggestions	0.09	0.047	0.457
DDD				
Information Exchange	Seeking info from all sources	0.147	0.136	1.454
	Passing info before asked	-0.002	0.008	-0.094
	Responding to request for info	-0.088	0.139	-0.845
Supportive Behavior	Correcting team errors	-0.208	0.617	-0.095
	Providing or requesting backup when needed	0.073	0.475	0.043
Leadership Initiative	Providing guidance or suggestions	-0.216	0.086	-0.588*

Table 4: Regression Analysis for TDT sub-dimensions. *Note*: Some sub-dimensions not analyzed due to lack of data points.

* p<.05

Ultimately, this current research, while only a first step, contributes to the larger body of research on teams and team processes. Specifically, the current study supports the theory that teams have both similarities and differences (i.e., in teamwork behaviors) based on their characteristics (e.g., distribution, interdependency). This research has implications in the design of future training. For example, distributed team training may need to target different teamwork behaviors (e.g., training that targets leadership initiatives: providing

guidance or suggestions) than training for co-located teams (e.g., training that targets supportive behaviors: passing information before asked). However, further testing is needed to determine the usefulness of such training.

Acknowledgement

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

References

Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors, 40(4),* 672-679.

Cannon-Bowers, J. A. & Salas, E. (1990). *Cognitive Psychology and Team Training: Training Shared Mental Models.* Paper presented at the 5th Annual Conference of the Society of Industrial and Organizational Psychology, Miami, FL.

Cohen, S. G. & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management, 23(3),* 239-290.

Driskell, J. E. & Salas, E. (1992). Collective behavior and team performance. *Human Factors, 34,* 277-288.

Fiedler, F.E., Potter, E.H., Zais, M.M., & Knowlton, W.A. (1979). Organizational Stress and the use and misuse of managerial intelligence and experience. *Journal of Applied Psychology, 64(6),* 635-647.

Griffith, J. (1997). Test of a model incorporating stress, strain, and disintegration in

the cohesion-performance relation. *Journal of Applied Social Psychology, 27(17),* 1489-1526.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31(5),* 519-537.

Hollenbeck, J.R., Sego, D.J., Ilgen, D.R., Major, D.A., Hedlund, J., and Phillips, J. (1997). Team decision making accuracy under difficult conditions: Construct validation of potential manipulations and measures using the TIDE² simulation. In M. Brannick, E. Salas and C. Prince (eds.), New Directions in Team Measurement, pp. 111-136. San Francisco: Jossey Bass.

Kleinman, D., Pattipati, K. Luh, P., and Serfaty, D. (1992). Mathematical models of team performance: A distributed decision-making approach. In R. Swezey and E. Salas, (Eds.)

Kozlowski, SWJ (1996, March). TEAMS / TANDEM: An experimental platform for examining skill acquisition, adaptability, and effectiveness at individual and team levels of analysis. Presented at the Personnel Human Resource Research Group, University of Florida, Gainesville, FL.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., & Salas, E. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85(2),* 273-283.

Milgram, N.A., Orenstein, R., & Zafrir, E. (1989). Stressors, personal resources, and social supports in military performance during wartime. *Military Psychology, 1(4),* 185-199.

Minionis, D. P. (1994). *Enhancing team performance in adverse conditions: The role of shared mental models and team training on an interdependent task.* Unpublished doctoral dissertation, George Mason University, Fairfax, VA.

Orasanu, J. (1990). Shared Mental Models and Crew Decision Making. Cognitive Science Laboratory Report #46. Princeton, NJ: Princeton University.

Saavedra. R., Earley, P. C., and Van Dyne, L. (1993). Complex interdependence in task performing groups. *Journal of Applied Psychology, 78(1),* 61-72.

Salas, E., Burke C. S. & Cannon-Bowers, J. A. (2000). Teamwork: emerging principles. *International Journal Review,* 2(4), 339-356.

Sanderson, P.M. and Fisher, C. 1994. Exploratory Sequential Data Analysis: Foundations. *Human-Computer Interaction,* 9, 251-317.

Smit, I. & Schabracq, M. (1997). Stress, performance, and organizational cultures. *International Journal of Stress Management, 4(4),* 275-289.

Smith-Jentsch, K. A.;, Zeisig, R. L., Acton, B. , & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers, and E. *Salas* (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271-297). Washington, DC, US: American Psychological Association.

Smith-Jentsch, K.A., Payne, S.C. & Johnston, J.H. (1996). Guided team self-correction: A methodology for enhancing experiential team training. In K.A. Smith-Jentsch (Chair), *When, how, and why does practice make perfect?* Paper presented at the 11th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Sundstrom, E. D. (1999). *Supporting work team effectiveness: Best management practices for fostering high performance.* San Francisco, CA: Jossey-Bass, Inc.

Sundstrom, E. McIntyre, M., Halfhill, T., and Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics, 4(1),* 41-67.

Thornton, R. C. (1992). *The effects of automation and task difficulty on crew coordination, workload, and performance.* Unpublished doctoral dissertation, Old Dominion University, Norfolk, VA.

Travillian, K. K., Volpe, C. E., Cannon-Bowers, J. A., & Salas, E. (1993). Cross-training highly interdependent teams: Effects on team process and team performance. *Proceedings of the 37th Annual Human Factors and Ergonomics Society Conference* (1243-1247). Santa Monica, CA: HFES Press.

Weaver, J. L., Bowers, C., & Salas, E. (2001). Stress and teams: Performance effects and interventions. In P A. Hancock and P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 83-106). Mahwah, NJ: Lawrence Erlbaum Associates.

Zaccaro, S., Parker, C., Marks, M., Burke, C. S., Higgins, J. M., & Perez, R. (1997, April). Team efficacy, communication, and performance: Implications for collective regulatory processes. In R. J. Klimoski (Chair), *Thinking and feeling while doing: Modeling team self-regulation.* Symposium conducted at the 12th Annual Meeting of the Society for Industrial And Organizational Psychology, St. Louis, MI.

EFFICACY OF STRESS EXPOSURE TRAINING ON TARGET ACQUISITION IN COMBAT SIMULATIONS

J.M, Ross, H.C. Ganey, & R.S. Broadway
University of Central Florida
Orlando, Florida

Transfer of training from simulators to real-world environments has recently come under investigation, as the generalizability of task-specific training has come into question. New hypotheses recommend that, to ensure effective performance under stress in real-world environments, one should supplement skill-based training with Stress Exposure Training (SET). Stress Exposure Training has further benefits in that it may serve as a more generalizable form of training and transfer across tasks and stressors. The impact of improving performance and reducing perceived stress and workload is of vital importance to many military operations, especially in high technology and high workload situations such as Landwarrior or Unmanned Arial Vehicles (UAVs; Mouloua, Hancock, & Gilson, 2003), in which mistakes are costly in terms of economics as well as life. In this paper the limits of SET transfer between laboratory training and field performance are investigated in regards to simulated combat target identification tasks.

INTRODUCTION

Simulation and training are often used to prepare individuals (e.g., pilots, soldiers) for performance in extreme environments. An assumption of man and training protocols is that the skills individuals develop in training transfer to real-world situations (see Dennis & Harris, 1998; Lane & Tang, 2000; Thomson, Ampofo-Boateng, Pitcairn, & Grieve, 1992). Training performance benefits are based on the premise that practicing real-world tasks in a simulator increase user familiarity with performing the task, provides users with feedback on performance, and allows one to practice skills in a consequence free environment. However, traditional task-based training programs are limited in their generalizability from training simulators to the real-world by domain specificity and altered decisional environments.

Domain Specificity

Domain specificity is defined in that skills acquired in training do not readily transfer to different tasks and situations (Sims & Mayer, 2002). This is critical due to the fact that situations in the real-world are highly complex and dynamic, making it virtually impossible to anticipate all potential accident scenarios that may occur. An example may be training a student pilot to respond to an aircraft engine failure, and then testing him or her on a loss of air cabin pressure. Unless training transfers across tasks, the scope of simulation training can be severely limited.

Altered Decisional Environment

The second facet that limits transfer of simulation training is a change in the decisional environment. That is, real-world situational pressures and stresses can change the decisional environment and limit the generalizability of skills obtained from simulation training (See Table 1). In a study by Zakay and Wooler (1984) participants who received skill training performed more effectively, but performance benefits disappeared when participants performed under time pressure. Thus a task-specific view of simulator training ignores emotional, behavioral, and physiological states induced by stressful real-world conditions, limiting transferability to real-world environments (Rachman, 1983; Driskell & Salas, 1991; Combs & Taylor, 1952; Mathews & Canon, 1975). Hence, training methods should be designed to consider the pressures that individuals would be performing under in real-world situations.

Stress Exposure Training

One method of increasing the generalizability of simulator-based training is through supplementing traditional task skill-based training with Stress Exposure Training (SET; Klepac, Hauge, Dowling, & McDonald, 1981; Saunders, 1993; Saunders, Driskell, Johnston, Salas, 1996). Considerable evidence has shown that stressors cause decrements in attitudes and performance related to emotional, behavioral, or physiological outcomes (See Table 1; Greenhaus & Parasuraman, 1987); to counter this, Stress Exposure Training trains

Table 1.
Emotional, Behavioral, and Physiological Effects of Stress on Human Performance. Adapted from Driskell & Johnston (1998).

Effects of Stress on Human Performance
• Physiological Changes
o Quickened Heartbeat
o Labored Breathing
o Trembling (Rachman, 1983)
• Emotional Reactions
o Fear, anxiety, & frustration (Driskell & Salas, 1991)
o Motivational losses (Innes & Allnutt, 1967)
• Cognitive Effects
o Narrowed attention (Combs & Taylor, 1952; Easterbrook, 1959)
o Decreased Search Behavior (Streufert & Streufert, 1981)
o Longer reaction time to peripheral cues and decreased vigilance (Wachtel, 1968)
o Degraded problem solving (Yamamoto, 1984)
o Performance rigidity (Staw, Sandelands, & Dutton, 1981)
• Changes in Social Behavior
o Loss of team perspective (Driskell, Salas, & Johnston, 2000)
o Decrease in prosocial behaviors (Mathews & Canon, 1975)

individuals on factors that are more generalizable than task skills, such as controlling perceived stress and workload by providing intervention information on stressors, sensory information on physiological and emotional reactions to stress, and procedural information on how stress affects task performance. These three phases of stress exposure training are further discussed in the following section.

The Three Phases of Stress Exposure Training

Phase 1. The first phase in stress exposure training typically involves discussion of common reactions people have to specific stressors they encounter, with the objective to help the individual understand that they will be able to stop the negative thoughts and behaviors that contribute to stress. It is expected that as the individual's anxiety is reduced, perceived control increases, and consequently ones efficacy in dealing with problem stressors is improved (Johnston & Cannon-Bowers, 1996). Interestingly, violations of expectations have been found to be universally detrimental to performance, even when participants' expectations are violated such that their situation is better than expected (Keinan & Friedland, 1996). The effect of credible preparatory information on performance has been found to result in reduced subjective stress, enhanced task confidence, and fewer errors (Inzana, Driskell, Salas, & Johnston, 1996).

Phase 2. The second phase focuses on learning stress coping skills through practice and feedback. The objective is to train the individual to maintain an awareness of stress reactions in order to invoke appropriate skills to reduce stress. The act of maintaining an awareness of thoughts and actions is referred to as "metacognition" (Glaser & Bassok, 1989). Therefore, metacognitive skills; such as, thought restructuring (replacing negative thoughts and reactions that are triggered by a stressor with positive coping thoughts and reactions), problem solving (reduce task performance errors), & physiological control (involves deep breathing and muscle relaxation methods to calm physiological reactions to stressful encounters) support the execution of competent performance.

Phase 3. The third phase involves practicing the coping skills in a setting that simulates or reproduces the problem stressors. Ways of coping with stress are often classified into three areas: task-focused responses, emotion-focused responses, and avoidant-focused responses (Lazarus & Folkman, 1984). Task-focused coping circumvents negative affect through cognitive and behavioral activity that minimizes or modifies the stressor. In this case the individual focuses on dealing with the task and develops strategies aimed at accomplishing the goals of the task (Thropp, Szalma, Ross, & Hancock, 2003). Emotion-focused coping consists of attempts to alleviate or eliminate emotions elicited by a stressor directly, with little attention paid to the characteristics of the situation or to the nature of the threat it poses (Keinan & Friedland, 1996). While, emotion-focused coping has been found to be effective in preparing individuals to cope with stressors that they cannot control and which are brief in duration, it has proved less effective/relevant for training people to withstand and cope with combat stressors. This more prolonged and controllable situations requires mixed or problem-focused coping. The third coping strategy, avoidant coping, involves diverting one's attention from the stressful situation (see also Cox & Ferguson, 1991). Expected outcomes from this phase are reduced anxiety, increased perceived efficacy toward performance, and improved cognitive and psychomotor performance. Further research involves examining phased-training procedures that divide task acquisition without stress exposure from practice of newly acquired skills under stress. Other methods of skill practice under stress include graduated-intensity training (Keinan & Friedland, 1996).

Presence and Immersion

Presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another. Immersion, on a related

note, is characterized by the sense of perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences (Witmer & Singer, 1998). With virtual environments one can manipulate the degree to which individuals experience presence and immersion through manipulating the computer-generated world (i.e., depth of simulation). Users are active seekers of information, who the more or less they perceive themselves to be inside the computer-generated world will impact the degree of meaningfulness of the simulation. Assuming that learning improves when the user is an integral part of the stimulus flow, and that meaningfulness and active control over a user's experiences aids learning, then immersive environments likely are better training tools than standard computer-based training environments (Witmer & Singer, 1998). That is factors increasing perceived presence would increase learning and performance of task relevant information.

Goal of Current Study

Driskell, Johnston, & Salas (2001) found that such stress training was generalizable across both novel stressors and novel computer based tasks. The primary goal for the current study is to extend the study of Driskell, Johnston and Salas by examining the generalization of Stress Exposure Training between laboratory training and field performance in a combat target acquisition task. The secondary goal is to discover how deep the presence (feeling of being in one place, while physically in another) and immersion (feeling of being in the simulation) felt during the simulation significantly modify the ability of operators to acquire targets. Additionally, we will look at how different presentation methods facilitate or inhibit the effects of SET to real-world situations when training individuals to control their emotional, behavioral, and physiological states in a combat situation. This last interactive effect is of particular importance, since it is hypothesized that the magnitude of improvement of performance from SET will be mediated by degree of presence.

METHOD

Experimental Participants

The participants in this experiment will be recruited from the University of Central Florida (UCF), and will receive course credit for their participation. Participants will be randomly assigned to one of the six conditions: Stress Training Virtual Reality (VR; High Presence), Control VR, Stress Training PC (PC; Low Presence), Control PC, Stress Training Mixed Reality (MR; Med Presence), or Control MR.

Experimental Design

To test our hypotheses, we will compare SET in three environments using Military Operations in Urbanized Terrain (MOUT) scenarios created by the UCF Institute of Simulation and Training (IST). These environments, in ascending level of presence and immersion, will include a PC-based Graphical User Interface (GUI; low presence and low immersion), a full Virtual Environment (VE; low presence and high immersion), and a Mixed Reality experience where some elements of the simulation are physically real and others are projected virtually (MR; high presence and high immersion). Using a 2 (training condition) x 3 (display) x 2 (pre/post) mixed model design, with repeated measures on the last factor, participants will be randomly assigned to Stress Exposure Training or a control group and to one of the three possible training environments (GUI, VE, or MR; 8 subjects per cell, N=48). Field performance will be evaluated in a field MOUT environment using an M-4 rifle, with DMX laser targeting, target acquisition and firing task before and after the simulation. The aim of this study is to examine whether transfer of SET principles from training to the operational environment will occur, and how presence and immersion facilitate or inhibit this transfer.

Experimental Tasks

During all tasks noise stress of 85dBA context relevant ambient noise will be played over the MOUT sound system. Noise stress will be manipulated by frequency and content. Each participant will participate in three tasks: a pre-test, a simulation session, and a post-test. The pre-test and post-test will be 8 minute in duration and will both be conducted at moderate noise stress. During these tasks participants will be instructed to complete a target acquisition task similar to those employed during military and police training (Stafford, Oron-Gilad, Szalma, & Hancock, 2004). Pre- and post testing will occur in an indoor MOUT environment employing pop-up targets. Order of target presentation will be counterbalanced between pre- and post-tests. The simulated sessions will be conducted in the same indoor MOUT system using a HMD and virtual targets. While in the simulated environment, participants will be asked at intervals to perform the simulated combat task for a total of 14 minutes.

SET Protocol

The treatment will be based on the 3-phased SET training protocol, where one group receives the SET training and one group receives an equivalent control condition. The SET training protocol is based off of the Attentional Focusing Training protocol developed by

Singer and colleagues (1991), and involves a preparatory handout describing the stressor the participant will be experiencing and typical reactions to that stressor, followed by a 15 minute video, and then graded intensity practice and review. The video is comprised of four sections: Relaxation, Visualization, Focusing Attention, and Refocusing attention after distraction. As a manipulation check after each section the video will be stopped and participants will be given a brief quiz on the content. The final section of graded practice and review involves 2 minutes practice in the simulated environment at low noise stress, followed by a review of skills, 4 minutes simulator practice at moderate noise stress, review of skills, and 8 minutes of simulator practice in moderate to high noise stress.

The group that does not receive the SET will be presented with general military information in the same formats and for an equivalent period of time that those in the other condition receive treatment. The final section of graded practice and review will be 14 continuous minutes of moderate intensity noise stress.

Dependent Measures

DSSQ. The Dundee Stress State Questionnaire (DSSQ; Matthews, Joyner, Gilliland, Campbell, Huggins, & Falconer, 1999) is a 104-item multidimensional self-report instrument for assessing transient states. Specifically, the DSSQ is made up of eleven sub-scales measuring Energetic Arousal, Tense Arousal, Hedonic Tone, Motivation, Self-focused Attention, Self-Esteem, Concentration, Confidence and Control, Task Relevant Interference, and Task Irrelevant Interference. These ten scales load onto three secondary scales: *Distress*, *Worry*, and *Task-Engagement*.

NASA-TLX. The NASA-Task Load Index (Hart & Staveland, 1984) was developed to assess the manner in which workload was perceived by operators in a system. It has six different axes: mental demand, physical demand, temporal demand, effort, performance, and frustration.

Presence Questionnaires. The Presence Questionnaire (PQ) and the Immersive Tendencies Questionnaire (ITQ) was developed by Witmer and Singer (1998) to quantify the amount of presence that a person feels while in a simulated environment.

Accuracy and Time to Target. This measure establishes task performance. Time to target will be acquired via a motion tracking system in the simulated MOUT site. Accuracy will be measured by looking at the length of the line between the center of the target and the point at which the weapon was pointed when the trigger was pulled.

RESULTS

Due to constraints of personnel and facilities results are pending as of the time of printing. For a full copy of results please contact the primary author.

DISCUSSION

To optimize the training and real-world application of simulator-based exercises it would be prudent to investigate further the generalizability of stress training across mediums as proposed in this study. This research proposal is based on broad foundation of past work in training for stress and is the next logical step in development of such methods. More generalizable simulator-based training methods will improve performance and decrease stress and workload levels across stressors and tasks. An outcome of this study would be recommendations for the improvement in the quality of training using simulators in many diverse areas including military training, aviation, and other settings in which effective performance under stress is required.

ACKNOWLEDGMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621 and facilitated by the Army Research Laboratory DAAD19-01-C-0065. P.A. Hancock, *Principal Investigator*. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors thank Dr. Sherry Tove, Dr. Elmar Schmeisser, Dr. John Lockett III, and Dr. Mike Drillings for providing administrative and technical direction for the Grant. The assistance of the Psychology Department at University of Central Florida is also gratefully acknowledged.

REFERENCES

Combs, A.W., & Taylor, C. (1952). The effect of the perception of mild degrees of threat on performance. *Journal of Abnormal and Social Psychology*, **47**, 420-424.

Cox, T., & Ferguson, E. (1991). Individual Differences, Stress, and Coping. In C.L. Cooper & R. Payne (Eds.), *Personality and stress: Individual differences in the stress process* (pp. 7-30). Oxford, England: John Wiley & Sons.

Dennis, K.A., & Harris, D. (1998). Computer-based simulation as an adjunct to ab initio flight training. *International Journal of Aviation Psychology*, **8**(3), 261-276.

Driskell, J.E., & Johnston, J.H. (1998). Stress Exposure Training. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 191-217). Washington D.C.: American Psychological Association.

Driskell, J.E., Johnston, J.H., & Salas, E. (2001). Does stress training generalize to novel settings? *Human Factors,*

43(1), 99-110. Mahwah, N.J.: Lawrence Erlbaum Associates.

Driskell, J.E., Salas, E, & Johnston, J.H. (2000). Does stress lead to a loss of team perspective. *Human Performance in Extreme Environments,* **5**(1), 69-76.

Driskell, J.E., & Salas, E. (1991). Overcoming the Effects of Stress on Military Performance: Human Factors, Training, and Selection Strategies. In G. Reuven & A.D. Mangelsdorff (Eds.), *Handbook of military psychology* (pp. 183-193). NewYork, NY: John Wiley & Sons.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review,* **66,** 183-201.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review Psychology, 40,* 631-666.

Greenhaus, J.H., & Parasuraman, R (1987). A work/nonwork interactive perspective of stress and its consequences. Special issue: Job stress: From theory to suggestion. *Journal of Organizational Behavior Management,* **8**, 37-59.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp.139-183). Amsterdam: North-Holland.

Innes, D.R., & Allnutt, M.F. (1967). Performance measurement in unusual environments (IAM Technical Memorandum No. 298). Farnborough, England: RAF Institute of Aviation Medicine.

Inzana, C.M., Driskell, J.E., Salas, E., & Johnston, J.H. (1996). Effects of preparatory information on enhancing performance under stress. *Journal of Applied Psychology, 81(4),* 429-435.

Johnston, J.H., & Cannon-Bowers, J.A. (1996). Training for Stress Exposure. In E. Salas (Eds.), *Stress and human performance* (pp. 223-256). Hillsdale, NJ: Erlbaum.

Keinan, G., & Friedland, N. (1996). Training Effective Performance Under Stress: Queries, Dilemmas, and Possible Solutions. In E. Salas (Eds.), *Stress and human performance* (pp. 257-277). Hillsdale, NJ: Erlbaum.

Klepac, R.K., Hauge, G., Dowling, J., & McDonald, M. (1981). Direct and generalized effects of three components of stress inoculation for increased pain tolerance. *Behavior Therapy,* **12**(3), 417-424.

Lane, D.M., & Tang, Z. (2000). Effectiveness of simulation training on transfer of statistical concepts. *Journal of Educational Computing Research*, **22**(4), 383-396.

Lazarus, R.S., & Folkman, S. (1984). *Stress appraisal and coping*. New York: Springer.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (vol 7), (pp. 335-350).

Mathews, K.E., & Canon, L.K. (1975). Environmental noise level as a determinant of helping behavior. *Journal of Personality and Social Psychology*, **32**, 571-577.

Mouloua, M., Gilson, R., & Hancock, P.A., (2003). Human-centered designs of unmanned aerial vehicles. *Ergonomics in Design,* **11**(1), 6-11.

Rachman, S. (1983). Fear and courage among military bomb-disposal operators. *Advances in Behaviour Research and Therapy*, **4**(3), 175.

Saunders, T.L. (1993). *The effect of stress inoculation training on anxiety and performance.* Unpublished master's thesis, University of Central Florida, Orlando, Florida, United States.

Saunders, T., Driskell, J.E., Johnston, J.H., & Salas, E. (1996). The effect of stress inoculation training on anxiety and performance. *Journal of Occupational Health Psychology,* **1**(2), 170-186.

Sims, V.K., & Mayer, R.E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology,* **16**(1), 97-115.

Singer, R. N., Cauraugh, J. H., Murphey, M., Chen, D., & Lidor, R. (1991). Attentional control, distractors, and motor performance. Human Performance, 4, 55-69.

Stafford, S.C., Oron-Gilad, T, Szalma, J.L., & Hancock, P.A. (2004). Time perception, shooting performance, and perceived workload in a police night-training shooting exercise. Manuscript submitted for publication.

Staw, R.M., Sandelands, L.E., & Dutton, J.E. (1981). Threat-rigidity effects in organizational behavior: A multi-level analysis. *Administrative Science Quarterly,* **26,** 501-524.

Streufert, S., & Streufert, S.C. (1981). Stress and information search in complex decision making: Effects of load and time urgency (Tech. Rep. No. 4). Arlington, VA: Office of Naval Research **.**

Thompson, J.A., Ampofo-Boateng, K., Pitcairn, T.K., & Grieve, R. (1992). Behavioural group training of children to find safe routes to cross the road. *British Journal of Educational Psychology,* **62**(2), 173-183.

Thropp, J.E., Szalma, J.L., Ross, J.M., & Hancock, P.A. (2003). Individual differences in dispositional pessimism, stress, and coping as a function of task type. *Proceedings of the Human Factors and Ergonomics Society, USA, 47,* 1073-1077.

Wachtel, P.L. (1968). Anxiety, attention, and coping with threat. *Journal of Abnormal Psychology,* **73,** 137-143.

Witmer B.G., & Singer, M.J. (1998) Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, **7**(3), 225-240.

Yamamoto, T. (1984). Human problem solving in a maze using computer graphics under an imaginary condition of "fire." *Japanese Journal of Psychology,* **55,** 43-47.

Zakay, D. & Wooler, S. (1984). Time pressure, training and decision effectiveness. *Ergonomics*, **27**(3), 273-284.

STRESS EFFECTS ON SOLDIER PERFORMANCE

Organizer and Panel Chair
C. Shawn Burke
University of Central Florida
Orlando, FL

Panel Co-Chair
Peter Hancock
University of Central Florida,
Orlando, FL

Panelists

Peter Hancock
University of Central Florida
Orlando, FL

Joel Warm
University of Cincinnati
Cincinnati, OH

Jim Szalma
University of Central Florida
Orlando, FL

Shawn Burke
University of Central Florida
Orlando, FL

Tal Oron-Gilad
University of Central Florida
Orlando, FL

Wayne Harris
University of Central Florida
Orlando, FL

Moderator

Clint Bowers
University of Central Florida
Orlando, FL

The complexity of the modern battlefield requires that significant practice and training must be provided to ensure success in situations where failure is not an option. Performance within these environments occurs against a background of threat which ensures that acute and chronic forms of stress are a constant presence. However, there remain many questions to be answered regarding the impact of stress on soldier performance. The proposed panel is comprised of experts whom have generated work from disparate paradigmatic perspectives to study the issues related to stress and human performance within the laboratory as well as the field. The panel's goal is to present/discuss critical issues that confront those who wish to understand soldier performance under stress. Although we focus on the soldier, in principle, our work can apply to any individual who finds themselves faced with decision uncertainty in ambiguous situations in which flawed performance can have serious consequences.

INTRODUCTION

At the turn of the 21st Century our military forces face a change in the challenge. We no longer confront an 'obvious' enemy. In recent engagements, it has become progressively more difficult to identify the opposing forces. Combined with this problem has come a comparable change in the site of engagement. No longer fought in wide open spaces, modern conflict promises to occur in the close quarters of shattered urban areas in which the support we seek to provide our troops may be negated by local circumstances. Into this new milieu, we are also evermore injecting innovative technologies whose functionality is designed to assist our troops, but whose complexity can threaten to overwhelm our young soldiers.

While soldiers are being required to improve their combat readiness for conditions which are dynamic and difficult to predict it is possible to anticipate some situations. For example, medical personnel responding to casualties may be asked to carry an injured soldier on a litter for a considerable distance and then require that

same individual to perform fine motor tasks involved with medical treatment. In another venue, maintenance personnel may well have to conduct repairs requiring significant physical effort while exposed to threat and decision-making uncertainty. Yet we do not have a thorough understanding of the cognitive demands of physical effort in situations such as these. For example, how do the combined influences of physical demand and threat impact decision-making and movement control? Further, some projections have circumstances in which a small team of soldiers travel significant time and distance in an enclosed vehicle, only to deploy on a moments notice. How does such abrupt workload transition to an extremely high level of demand affect their capacity to respond? While, fortunately, we have some familiarity with such situations (Hancock, 1997; and see Huey & Wickens, 1993), much more information is needed to provide predictive models to help distinguish the general and individual effects of such transitions.

While the above represent field examples of stress and the potential impact on cognitive readiness there are also problems of stress, in command, control, and decision-making at both the individual and team levels. For example, at the team level it has been argued that team decision-making is a process requiring team members to "filter raw data, apply individual expertise, communicate relevant information, and often make recommendations to other members" (Cannon-Bowers, Salas, & Converse, 1993, p. 222). This creates a situation where individual task performance is a necessary, but not sufficient condition for effective team decision making. Specifically, the quality of the team's decision rests on their ability to exhibit appropriate teamwork in order to function as a coherent unit. On the battlefield, soldiers are confronted with many sources of stress which may affect their ability to make quality team decisions, not the least of which is the uncertainty and cognitive overload caused by the increase in information afforded to teams, both in quantity and complexity. Additionally task (e.g., distribution) and team characteristics (e.g., team type, team composition) may differentially impact the relationship between stress, teamwork processes, and the resulting performance. However, there exists very little guidance as to how to increase a team's cognitive ability to make quality team decisions when operating in stressful, complex environments.

This lack of understanding poses a problem as the increasing complexity of the modern battlefield requires that significant practice and training at all levels of operation be provided to ensure success in situations where uncertainty, abrupt workload transitions, and time pressure are a constant presence. However, fundamental issues regarding stress, workload, and fatigue on the modern electronic battlefield remain to be addressed. In addition, innovative ways in which such cognitive readiness can be promoted must also be considered.

THE PANELISTS

This panel session will draw on the knowledge and experiences of behavioral scientists that have been actively confronting the challenges of understanding the impact that stress has on soldiers. Each of the panelists has been engaged in work under a five-year programmatic effort funded by the Army Research Office. Specifically, the panelists are part of a Multi-University Research Initiative (MURI) whose goal is to examine the impact of stress on cognitive readiness and soldier performance on the electronic battlefield. The panelists represent a range of experts spanning multiple fields, including cognitive science, experimental psychology, physiological psychology, and social science. This diverse yet complementary group has served to provide the synergy necessary to address the fundamental issues surrounding this problem. We pull from the expertise on the proposed panel session to present critical issues that confront those who wish to understand soldier performance under stress. Although we focus on the soldier, in principle, our work can apply to any individual who finds themselves faced with decision uncertainty in ambiguous situations in which failing or flawed performance can have serious if not fatal consequences.

While the initial portion of our panel examines issues related to stress with regard to attention and vigilance, the later portion of our panel members will speak to the impact that stress has on team process and performance, methods that can facilitate training soldiers to better respond to stress whether they are embedded within teams or solely operating as individuals, and finally lessons about assessment of soldier stress that have been learned in the field.

As can be seen from the individual summaries (see below), panel members represent a wide range of expertise both in terms of approaches, focus, methods, and assessment strategies. The views presented on the proposed panel also cross a range that is comprised of: laboratory work that is strict in its methodologies, to work that combines aspects of laboratory and field assessments, to the other end of the continuum, collection of data within field environments. It will be Dr. Bowers role, in his capacity as panel moderator, to focus the discussions to highlight the commonalities as well as distinct challenges associated with the approaches presented within the proposed panel. Given the wide diversity of panel members' perspectives it is

our belief that this panel offers a unique opportunity to highlight the critical issues for those interested in examining the impact of stress on soldier performance.

ABSTRACTS OF EACH PANELIST'S OPENING COMMENTS

Peter Hancock
University of Central Florida, Orlando, FL

Stress and the Role of Today's Soldier. Recent International events have found the soldiers of the United States and the coalition placed in an invidious situation. Trained for various forms of combat, they now find themselves in the role of peacekeepers in cultures that are strange and alien to them. Dr. Hancock will argue that this activity represents at least the same if not a greater degree of stress since the goal of the task is much less explicit than is expressed in their combat activities. As such, our models of stress and performance and how response readiness and response capacity varies, are as applicable to this new role as they are to the traditional pursuits of the military. In particular, he will focus on the problem of workload transitions or the more commonly termed 'hours of boredom and moments of terror.' These periods of transition are related to our refined model in order to understand soldier response in uncertain and volatile conditions.

Joel S. Warm
University of Cincinnati, Cincinnati, OH

External vigilance: Necessary but stressful. Dr. Warm will speak to work that he is conducting with colleagues Matthews and Parasuraman. This reflects a discussion of the stress that vigilance can impart on a soldier's performance. Specifically, research indicates that the stress imposed by vigilance work may not be due to the relative boredom that comes from vigilance tasks, but in fact that stress may be caused by the fact that these tasks are very resource demanding. These researchers thereby present a different picture than that of traditional vigilance work. In order to detect enemy threats, modern warfare systems require soldiers to maintain sustained attention or vigilance to a variety of displays ranging from computer consoles to warning systems on combat helmets. Hence, it is necessary to understand the implications of this requirement for military personnel. According to a recent model by Robertson and his associates (Robertson, et al., 1997) the repetitive nature of vigilance assignments leads monitors to a thoughtless, routinized, approach in which effortful attention is withdrawn from the task at hand and consequently, to failures of signal detection. To the contrary, our research, funded by MURI, indicates that vigilance tasks are resource demanding and highly stressful. These findings are based upon elevated workload ratings on the NASA-TLX, high Distress scores on the Dundee Stress State Questionnaire, and measurements of cerebral blood flow indicating an exhaustion of information-processing resources.

Jim Szalma
University of Central Florida, Orlando, FL

Examination of attentional mechanisms underlying stress and performance. Dr. Jim Szalma will speak to the lessons that have been learned through empirical work on the impact that stress has on attentional mechanisms which comprise soldier performance. The Hancock and Warm (1989) model describing stress and performance articulated two primary dimensions along which task parameters impact an individual's adaptation to stress. These are information structure, primarily spatial in nature, and information rate, reflecting the temporal elements of tasks. Our research has focused on testing hypotheses regarding attention to spatial and temporal task elements, and the degree to which these co-vary. Preliminary results indicate that stress does not impact these dimensions equivalently. In a related line of research, we are testing elements of Fuzzy Signal Detection (FSDT; Parasuraman, Masalonis, & Hancock, 2000) for application to understanding detection performance under stress in complex real-world domains. The goal for the current work is to complete basic experimentation to test specific parameters of the model, with the aim of applying these results to real-world threat detection tasks laboratory and field studies.

C. Shawn Burke
University of Central Florida, Orlando, FL

Team performance under stress. Dr. Burke will examine the reaction of teams in the face of stress and will describe the approach and initial results taken in examining the impact of stress on different team 'types'. Military forces are facing situations where teams are required to operate in environments where stress is a defining characteristic and problems are ill-defined. While there has been a wealth of information collected on how stress impacts individual performance, less has been collected on how it impacts teams. Moreover, in general, team research has investigated global aspects of teams, failing to probe deeply into these complex entities. Results are consequently often applied as if all teams are created equal despite practical and theoretical evidence that all teams are not created equal (Sundstrom,

McIntyre, Halfhill, & Richards, 2000), nor do they have equal amounts of interdependency (see Saavedra, Earley, & Van Dyne, 1993). We argue that it is time to return to basics and investigate the defining characteristics of teams. Our research has focused on the implementation of a paradigm that allows the examination of how stress impact different team types.

Tal Oron-Gilad
University of Central Florida, Orlando, FL

Stress and performance in virtual environments. Dr. Tal Oron-Gilad will serve as a bridge across all panel members in that she will speak to her experiences with how virtual environments can be used to both study the impact of stress as well as how virtual environments can facilitate training. In Virtual environments, the display is close or 'proximal' to the individual and completely surrounds them in the fashion of the real-world. This technology is founded upon the same conception as high-fidelity simulation, except that such simulation relies upon some real-world components such as a cab or flight deck, while VR does not. As a result, VR offers the potential opportunity to provide portable, inexpensive, easily manipulable, surrogate worlds used for training and educational purposes. Capturing stress elements of real world tasks and environmental stressors and incorporating them in VR is critical for designing conditions suitable for efficient learning and improving performance, and for learners to flexibly adapt under multiple degrees of task load and complexity.

Wayne Harris
University of Central Florida, Orlando, FL

Field assessment of individual soldiers in stressful operational conditions. Finally, moving into the field Dr. Wayne Harris will describe his experiences in collecting field data concerning individual responses to stressful operations. Dr. Harris will specifically focus on observations related to the monitoring of soldier's whom have failed the battery of psychological tests of cognitive and psychomotor functioning and what this means in terms of the assessment battery. I here report on the range of individual differences that we have observed in soldier's reactions to the stress of demanding operational field conditions. Over the past three years, we have collected data from soldiers as the SERE School's and at field exercises at Ft. Chaffee and Ft, Bragg. Using hand-held devices, we have administered a battery of psychological tests of cognitive and psychomotor functioning which have provided diagnostic of performance state. We have found that of all tests, simple Reaction Time (RT) provides the fastest and most reliable portraiture. The focus of my talk will be on our experiences of monitoring failing soldier's, some of whom have had to be Med-Evac'ed from the field environment. Data from these individuals show the diagnostic power of recording response variability rather than relying on a single data point or on a collective mean. The interpretation of this observation will be discussed.

ACKNOWLEDGEMENTS

The work of panel members was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621, P.A. Hancock Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect official US Army policy.

REFERENCES

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision-making. In N. J. Castellan Jr. (Ed.), *Individual and group decision-making* (pp. 221-246). Hillsdale, NJ: Erlbaum.

Hancock, P.A. (1997). *Essays on the Future of Human-Machine Systems*. Banta: Eden Prairie: MN. (Second Printing, 1998, Third Printing, 1999, CD Release, 2000).

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519-537.

Huey, B. M., & Wickens, C. D. (1993) (Eds.). *Workload transition: Implications for individual and team performance*. Washington, DC: National Academy Press.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, 42(4), 636-659.

Saavedra, R., Earley, P. C., & Dyne, L. V. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology, 78*(1), 61-73.

Sundstrom, E., McIntyre, M., Halfhill, T., & Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990's and beyond. *Group Dynamics, 4*(1), 44-67.

THE IMPACT OF DISPOSITIONAL OPTIMISM AND PESSIMISM ON STRESS AS A FUNCTION OF PSYCHOPHYSICAL TASK CHARACTERISTICS

J.M. Ross[1], J.L. Szalma[2], and P.A. Hancock[12]
University of Central Florida[1] & Institute of Simulation and Training[2]
Orlando, Florida

In this study we tested the effect of individual differences in dispositional optimism and pessimism on target detection. Following proximal task dimensions set out by the Model of Maximal Adaptability, three tasks were employed in the current study, varying in their spatial and temporal emphasis. Performance as measured by self report of stress was examined in light of individual differences in attentional narrowing across the three task levels. Prior research indicates that optimism predicts increased *Task-Engagement*, but in this study there was a novel finding that higher levels of optimism predicted decreased *Task-Engagement* in spatial-dominant tasks. Optimism was also found to predict increased levels of post-task *Worry* in temporal-dominant tasks. However, expected relations between pessimism as post-task stress state were not observed. Although the results of the current study did not confirm the hypotheses regarding joint resource capacities for spatial and temporal task components, results did indicate that dispositional optimism does impact *Task-Engagement*, signifying a need for further research on the relation between this trait and participant stress. The results extend prior findings that the relationship between these traits and stress states may depend on the psychophysical characteristics of the target detection task employed.

INTRODUCTION

The operation of modern complex systems induces considerable stress in operators, who must then devote cognitive and physical resources to adapt to the stressor and maintain adequate performance levels. Design of systems that mitigate these effects require development of theoretical models articulating the mechanism of stress effects. Toward that end, Hancock and Warm (1989) described a dynamic model of stress and performance that explicitly recognized that *tasks themselves* represent a major proximal source of stress for operators. They further identified two task dimensions that influence adaptive processes and performance: 1) information structure, a *spatial characteristic*; and 2) information rate, a *temporal characteristic*. One possible mechanism by which stress impacts performance is via distortion of perceptions along these two dimensions, often manifested as attentional narrowing (Easterbrook, 1959). Evidence for the narrowing of attention under stress has, on the spatial dimension, been experimentally validated (e.g., see Cornsweet, 1969; Dirkin & Hancock, 1984). It has recently been proposed that the distortion of time perception results from the same mechanism as narrowing of spatial perception, and that therefore a joint distortion of spatial and temporal stimuli, as a result of declines in a common resource capacity for these dimensions, reflects a general effect of stress on perception of space-time (Hancock and Weaver, in press; Hancock, Szalma, & Weaver, 2002).

Recent evidence indicates that the stress experienced by participants is influenced by the observer's dispositional pessimism and optimism (Szalma, 2002; Thropp, Szalma, Ross, & Hancock, 2003). In addition, Thropp and her colleagues reported that these effects depend on the spatial and temporal properties of the task. The current study was designed to explore these effects by determining the effect of conjunctions of spatial and temporal discriminations and dispositional pessimism and optimism on detection performance and self reports of stress.

EXPERIMENTAL METHOD

Experimental Participants

Twenty-three female and twenty-three male undergraduates participated in this experiment (N=46). They ranged in age from 18 to 36 years old, with a mean age of 23.3 years (SD=4.2). All participants experienced each of the three tasks and were assigned at random to one of the six possible task order conditions. The order in which the tasks were completed was counterbalanced across participants to control for potential order effects.

Experimental Tasks

The three tasks which were employed in the present study were each seven minutes in duration. In each task, participants were required to discriminate a vertical line that varied along two of three dimensions: spatial (line length), temporal (stimulus duration), and luminance. Note that the number of features in each conjunction search was constant across the three tasks. Lines were presented on a background mask of 0.4mm diameter white circles on a black background (See Figure 1). The luminance of the white circles was 142.17 cd/m^2 and that of the black background was 0.95 cd/m^2. Ambient illumination in the room was 0.33 cd/m^2. The mask was adapted from one employed previously in a study by Temple, Warm, Dember, Jones, LaGrange, and Matthews (2000).

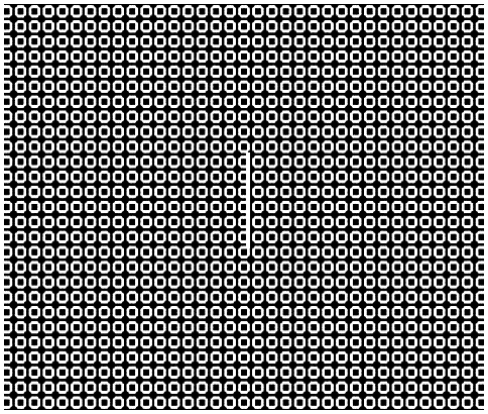


Figure 1. *Background mask with line stimulus. Note stimulus altered for illustrative purposes.*

The tasks differed according to spatial and temporal properties of target detection. The 'spatial-dominant' task required observers to detect a line of higher luminance and shorter length (3.3cm by 0.1cm). Non-targets were presentations of the line of longer length (3.5cm by 0.1cm) and higher luminance, or a lower luminance line of either length. Thus, the spatial-dominant task presented observers with a spatial discrimination regarding line length, while temporal properties were held constant. All stimuli appeared for 200ms. The 'temporal-dominant' task required a temporal discrimination instead of a spatial discrimination, and observers were required to report when they observed a line of higher luminance and shorter duration (165ms). Non-targets were presentations of a higher luminance line of longer duration (200ms), or a lower luminance line of either duration. In this condition, however, there was no spatial discrimination; stimuli were all lines of the longer length (3.5cm by 0.1cm). In the third task ('combined'), the temporal discrimination was combined with spatial discrimination. Thus, observers were required to detect a line of the shorter duration (165ms) and shorter length (3.3cm by 0.1cm). In this task luminance was held constant. Note that luminance was varied in the 'spatial-dominant' and 'temporal-dominant' tasks in order to maintain the same number of discrimination conjunctions across the three tasks.

The stimulus presentation rate was twenty-four events per minute with six signals per minute (signal probability=25%). The order of stimulus presentation was randomized for each participant. For each task, participants responded to a signal by pressing the spacebar key on the keyboard. Correct responses occurring within 2.5 seconds of the onset of a signal were recorded as Correct Detection while all other responses were recorded as False Alarms.

Questionnaires

Self-reports of stress were measured using the Dundee Stress State Questionnaire (DSSQ) a multi-dimensional instrument for assessing transient states associated with mood, arousal, and fatigue (Matthews et al., 2002; see also Matthews et al., 1999). The DSSQ is composed of eleven scales reflecting the observer's cognitive and emotional states. These include: Tense Arousal, Hedonic Tone, and Confidence and Control, reflecting the individual's degree of *Distress*; Energetic Arousal, Concentration, and two kinds of Motivation (Intrinsic and Success) which indicate the individual's degree of *Task-Engagement*; and Self-Esteem, Self-Focused Attention, and two forms of Cognitive Interference (Task-Irrelevant and Task-Related) which reflect the individual's degree of *Worry*.

Pessimism and optimism were measured using the Optimism/Pessimism Inventory (OPI; Dember, Martin, Hummer, Howe, & Melton, 1989; Dember, 200), a measure which treats these two traits as partially independent dimensions and provides separate scores for pessimism and optimism (See Hummer, Dember, Melton, & Schefft, 1992).

Experimental Procedure

Participants were instructed to complete the OPI prior to receiving any instructions regarding the tasks. Following completion of the OPI, the instructions for the first task were read to participants who then completed the pre-DSSQ, after which they began the first task. Upon completion of each task, participants were administered the post-DSSQ.

EXPERIMENTAL RESULTS

Performance

Regression analysis was employed to examine the effects of pessimism and optimism on performance and stress for each task. Separate regressions were computed for optimism and pessimism. Consistent with previous findings (Szalma, 2002; Thropp, et al., 2003), pessimism and optimism did not significantly impact overall performance, as measured by the proportion of Correct Detections and False Alarms (p>.05).

Self-Reported Pre-Task Stress State

Regression analysis on pre-task stress state scores results as a function of pessimism and optimism are shown in Tables 1 and 2. The results confirm prior research showing that pessimism predicted higher *Distress* and lower *Task-Engagement* and that optimism predicted lower *Distress* and higher *Task-Engagement* prior to task performance. Also consistent with the findings of Szalma (2002) was that neither trait was related to pre-task *Worry*.

Table 1.
Summary of Regression Analysis for Pessimism Predicting Pre-Task Stress State.

Variable	b	SE_b	β
Worry	.02	.02	.17
Distress	.05	.02	.32*
Task-Engagement	-.05	.02	-.41**

*p<.05, **p<.01*

Table 2.
Summary of Regression Analysis for Optimism Predicting Pre-Task Stress State.

Variable	b	SE_b	β
Worry	.02	.02	.16
Distress	-.05	.02	-.35*
Task-Engagement	.05	.02	.37*

p<.05

Self-Reported Post-Task Stress State

Pessimism. Results of regression analyses of each post-task stress state score for distress as a function of pessimism are shown in Table 3, analyses for post-task *Worry* and *Task-Engagement* were not significant (p>.05). However, contrary to the findings of Thropp et al. (2003), pessimism did not significantly impact stress state for any of the tasks in this study after pre-task state was taken into account. Hence, the spatial task-specific effects observed by Thropp et al. (2003), in which pessimism influenced *Task-Engagement* only in tasks with spatial uncertainty, did not extend to a case in which spatial discrimination is considered rather than spatial uncertainty.

Table 3.
Summary of Regression Analysis for Variables Predicting Post-Task Distress as a Function of Pessimism.

Variable	b	SE_b	β
Spatial Task	.07	.03	.33*
Temporal Task	.06	.03	.27
Combined Task	.05	.03	.24

p<.05

Optimism. For optimism, two sets of regressions were calculated: one in which pre-state was excluded from analysis and one in which pre-state was included in analysis. This was done so that the regressions could be compared to determine whether the influence of traits on post-task stress was mediated by pre-task state. Optimism significantly predicted increased post-task *Task-Engagement* in the temporal task (b=.04, SE_b=.02, β=.30, p<.05), *Distress* across two of the three task dimensions (See Table 4), and *Worry* in the temporal dimension (b=.05, SE_b=.02, β=.31, p<.05). However, with the exception of a trend toward significance with Optimism predicting *Worry* in the temporal-dominant task (b=.03, SE_b=.02, β=.21, p=.067), these regressions were not significant when the corresponding pre-task stress state was entered first into the regression equation.

Table 4.
Summary of Regression Analysis for Variables Predicting Post-Task Distress as a Function of Optimism

Variable	b	SE_b	β
Spatial Task	-.07	.03	-.34*
Temporal Task	-.07	.03	-.30*
Combined Task	-.06	.03	-.28

p<.05

In analyzing results after pre-task state was accounted for in the model, optimism was found to significantly predict decreased *Task-Engagement*, $F(2,43)=11.69$, p<.001, R^2=.35, $\Delta R^2_{optimism}$=.08, but this main effect was restricted to the spatial-dominant task. Optimism did not

significantly predict post-task *Task-Engagement* in the temporal-dominant or the combined tasks. Note that task was a within-subjects variable, so the individuals who reported decreased engagement after the spatial-dominant task did not report a similar change after the temporal-dominant task. In addition, preliminary analyses indicated that task order did not significantly impact post-task *Task-Engagement*.

DISCUSSION

Although the results of the current study did not confirm the hypotheses regarding joint resource capacities for spatial and temporal task components, results indicate that while pessimism predicted post-task *Distress* and optimism predicted decreased post-task *Distress*, this effect was non-significant when the pre-task stress state was entered first into the regression. Thus, as in the study by Szalma (2002), pessimism and optimism exert their effects on task-induced mood primarily by influencing the affective state of the individual prior to task performance.

After accounting for pre-task state, it emerged that dispositional optimism does impact one dimension of stress (*Task-Engagement*). However, as observed in previous research, the effects of this variable were task-dependent and emerged only when the pre-state was accounted for in the model. Further, the direction of the effect was unexpected, since in previous experiments increased optimism predicted increased post-task *Task-Engagement* rather than the novel finding of increased optimism predicting decreased post-task *Task-Engagement* observed herein (Szalma, 2002; Thropp, et. al., 2003). This effect was not a general one since it was restricted to the spatial task (recall that in this experiment task was a within subject variable). These results confirm the dominance of spatial characteristics observed by Thropp et al. (2003), suggesting that the spatial dimension may be more salient than the temporal dimension. Future research will investigate the degree to which the apparent dominance of spatial characteristics over temporal characteristics generalizes to other task domains. In addition, further research is needed to determine what aspects of a spatial task induces decreased *Task-Engagement* among individuals high in optimism, and why this effect emerges only after the pre-state is accounted for in the regression model.

ACKNOWLEDGMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. P.A. Hancock, *Principal Investigator.* This study was also facilitated by the Department of Defense through the Army Research Laboratory under grant DAAD19-01-C-0065. P.A. Hancock and J.L. Szalma *Co-Principle Investigators.* The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors thank Dr. Sherry Tove, Dr. Elmar Schmeisser, Dr. John Lockett III, and Dr. Mike Drillings for providing administrative and technical direction for the Grant. The assistance of the Psychology Department at University of Central Florida is also gratefully acknowledged.

REFERENCES

Cornsweet, D.M., (1969). Use of cues in the visual periphery under conditions of arousal. *Journal of Experimental Psychology,* ***80***(1), 14-18.

Dember, W.N., Martin, S.H., Hummer, M.K., Howe, S.R., & Melton, R.S. (1989). The measurement of optimism and pessimism. *Current Psychology: Research and Reviews,* **8***,* 102-119.

Dirkin, G.R., & Hancock, P.A. (1984). Attentional narrowing to the visual periphery under temporal and acoustic stress. *Aviation, Space and Environmental Medicine*, ***55***, 457.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, ***66***, 183-201.

Hancock, P.A., Szalma, J.L., & Weaver (2002). The distortion of perceptual space-time under stress. *Proceedings of the 23rd Army Science Conference*, Orlando FL.

Hancock, P.A. & Warm, J.S. (1989) A dynamic model of stress and sustained attention. *Human Factors,* **31***,* 519-537.

Hancock, P.A., & Weaver, J.L. (in press). On time distortion under stress. *Theoretical Issues in Ergonomic Science*.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp.139-183). Amsterdam: North-Holland.

Hummer, M.K., Dember, W.N., Melton, R.S., & Schefft, B.K. (1992). On the partial independence of optimism and pessimism. Current Psychology: Research and Reviews, **11**, 37-50.

Matthews, G., & Campbell, S.E. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society*, **42,** 821-825.

Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., et al. (2002). Fundamental dimensions of subjective state in performance settings: *Task-Engagement*, distress, and *Worry*. *Emotion*, ***2****(4)*, 315-340.

Matthews, G. Joyner, L., Gilliland, K., Campbell, S. Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a

state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, and F. Ostendorf (Eds.), *Personality psychology in Europe* (vol 7), (pp. 335-350).

Szalma, J.L. (2002). Individual differences in the stress and workload of sustained attention. *Proceedings of the Human Factors and Ergonomics Society*, **46**, 1002-1006.

Temple, J.G., Warm, J.S., Dember, W.N., Jones, K.S., LaGrange, C.M., & Matthews, G., (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors, **42**(2)*, 183-194.

Thropp, J.E., Szalma, J.L., Ross, J.M., & Hancock, P.A. (2003). Individual differences in dispositional pessimism, stress, and coping as a function of task type. *Proceedings of the Human Factors and Ergonomics Society*, **47**, 1073-1077.

INDIVIDUAL DIFFERENCES RELATED TO SHOOTING PERFORMANCE, IN A POLICE NIGHT-TRAINING SHOOTING EXERCISE

S.C. Stafford, T. Oron-Gilad, J.L. Szalma, and P.A. Hancock
University of Central Florida, Orlando Florida

Increased understanding of the stress and performance effects of gun range drills with police officers is of importance for law enforcement trainers. Seventy-one police officers participated in routine night shooting drills over a two week period as part of regular training regimen. This paper examines the effects of two shooting range tasks on police officer's time perception, individual differences, and shooting performance. Theoretical considerations led to the hypothesis that performance changes under stress may be related to distortions in the perception of time. However, the results of this study do not provide strong evidence for nomothetic effects of time distortion. Rather, it is more likely a function of individual differences. Future studies will examine potential individual differences variables that influence the relation between time distortion under stress and task performance.

INTRODUCTION

Police officers are acutely aware of the dangers of policing, and patrol officers in particular are often at the greatest risk, since they are the first officers encountered by victims, suspects and witnesses. Patrol officers frequently operate in hostile environments, confronting variable levels of threat and danger. For instance, in 2002 55 law enforcement officers were shot to death and 58,066 officers were assaulted in the U.S. of the latter, 16,494 sustained serious injury (FBI, 2003). Officers are taught that a key line of defense against violent assault is their firearm, and the skill and judgment with which they use it. Therefore officers are required to complete routine firearms assessment and training programs, such as the one described in this study. There has been very little research on the factors that influence performance in these routine shooting exercises. In this study we assessed the perceived workload, shooting performance, and time perception of officers engaged in these exercises. Detailed understanding of the officers' performance and stress level in the training exercise can lead to improved training procedures and thereby enhance shooting performance in real-world threatening situations.

The way people estimate time can possibly be manipulated by stress state. As Hancock and Warm (1989) have noted, tasks themselves are often the proximal source of stress, and individuals cope with this stress by narrowing their attention to specific cues in the environment (Easterbrook, 1959). Hancock and Weaver (2004) have shown that such narrowing occurs along the temporal domain, resulting in distortions of perceived time. At a phenomenological level, individuals who experience 'time distortion' under stress report one of two forms of experience, in which times appears to speed up or slow down (see Fair, 1984). Both of these patterns represent effects of 'time in memory' that are evident in the memory for the event. In stress conditions sufficient to induce time distortion, the increase in event registration must be substantial but to a level yet to be determined by empirical investigation. Temporal distortions may occur anywhere that insufficient attention is directed to temporal cues (Hancock & Weaver, 2004). However, evidence regarding time distortion under stress has been largely anecdotal in nature.

Field evaluations are of great importance to the study of performance under stress because they provide a realistic setting for observation. The current study examines individual differences, performance and time perception of police officers in gun training exercises. Of particular interest is the sensitivity of the Adjective Checklist as developed by Goldberg (1992) to measures of shooting performance.

METHOD

Participants

Participants were 71 police officers (10 women and 61 men) from a police department in the southeastern United States. The average age was 37 years (range 22-56 years). Police officers are required to participate in the evening night shooting exercise as part of their regular training schedule. However, they volunteered to participate in the research while performing the shooting exercise. Their decision whether to participate in the study did not influence their performance evaluation by their instructors.

Instrumentation

Officers were equipped with their personal SIG (Schweizerische Industrie Gesellschaft) Sauer p226 9mm handgun. All officers used the same standard issued duty belt with identical magazine pouches. Each officer also carried a standard department issue Streamlight Stinger flashlight 7.4 inches long and weighing 10 oz. All officers were required to wear a protective vest, hearing protection (external muffs, or internal plugs) and clear eye protection. The ammunition used for the task is a special bullet designed to shatter on contact with metal targets so as to reduce the likeliness of ricochet.

Materials

Self-reports of stress were measured using the Dundee Stress State Questionnaire (DSSQ; Matthews et al., 2002; see also Matthews et al., 1999). The DSSQ is composed of eleven factor analytically determined scales, organized into three secondary factors, reflecting the observer's cognitive and emotional states. These include Tense Arousal, Hedonic Tone, and Confidence and Control, reflecting the individual's degree of Distress; Energetic Arousal, Concentration, and two kinds

of Motivation (Intrinsic and Success; see Matthews, Campbell, & Falconer, 2001), which indicate the individual's degree of Task-Engagement; and Self-Esteem, Self-Focused Attention, and two forms of Cognitive Interference (Task-Irrelevant and Task-Related) which reflect the individual's degree of Worry. Following the instructions for the exercise participants completed the pre-DSSQ, after which they began Task A. Upon completion of all tasks, participants were administered the post-DSSQ.

Personality differences were measured using the adjective checklist (ACL) developed by Goldberg (1992) to assess the "Big Five" factors of personality. Following the instructions for the exercise participants completed ACL.

Task Design

The night shooting exercise included four different shooting tasks (referred to as Tasks A-D). The tasks differed in complexity and in their settings however, the major difference between tasks A, B, C and Task D was that in Tasks A, B, and C officers were aiming at paper targets and therefore in the dark could not easily gauge their performance, while Task D was a feedback shooting task where the officers aimed toward metal targets where they could clearly hear whether they hit the target or not.. Tasks also build up in difficulty and the two more complex tasks were tasks C and D. For the purpose of this study, the order of the last two tasks varied among participants, hence the order of the tasks was either A, B, C, and D, or A, B, D, and C. However due to the non-intervening nature of the study we did not have any control on who will be in either one of the conditions and therefore can not claim homogeneity of groups.

The order of the tasks (Task C and Task D) was counterbalanced between participants. The tasks differed in complexity and in their settings however, the major difference between the tasks was that Task C was a non-feedback shooting task where the officers aimed toward paper targets (and in the dark could not easily gauge their performance). While Task D was a feedback shooting task where the officers aimed at metal targets and could immediately receive feedback regarding their performance as a result of the sound of the bullet hitting the metal target. In both tasks the participants were running from target to target and aiming at different targets following the instructions of the shooting range instructor.

Procedure

Officers attended the training session in groups of six to twelve each, but each officer performed each task separately. The training session took place in the dark at an outdoor police shooting range, and the total session time varied from approximately 60 minutes to 120 minutes depending on the number of officers in the session. All officers were given the same instructions for procedural and safety reasons. Additionally, the officers were told that they will be asked by the experimenter to estimate of the task length in minutes and seconds upon completion of each task (Task C and Task D). There was short break between tasks.

RESULTS

Performance Measures

Shooting performance was calculated separately for each task by calculating the percentage of hits from the total shots made. This measure is the most simplified measure of performance because within each task shots were made from different ranges, so shooting accuracy and difficulty varied even within a single task. However, the hit percentage provides a global indication of how well officers performed on each task.

As expected, hit rate varied among tasks where the highest rate was found in Task A which was a stationary warm up task (74%, SD=19%), followed by Task B (65%, SD=22%), Task D (63%, SD=15%) and Task C(57%, SD=20%). There were significant differences between task A and the other tasks, and between task B and Task C, but there were no significant differences between Task C and Task D and between Task D and Task B, as shown in Table 1. We also found correlations between Task A and Task B, Task A and Task C, and Task B and Task C but no correlation with task D which was a metal shooting target while the other three were paper shooting targets. This may indicate that although the task itself has changes and became significantly more complicated (which is reflected by the degradation in hit rate) overall, the paper shooting target tasks were "more of the same" while the metal target in Task D provided "something different". Changing the order of the tasks between Task C and Task D did not have any effect on the shooting performance.

Table 1 – Pairwise Comparisons and correlations of hit percentages among the four tasks

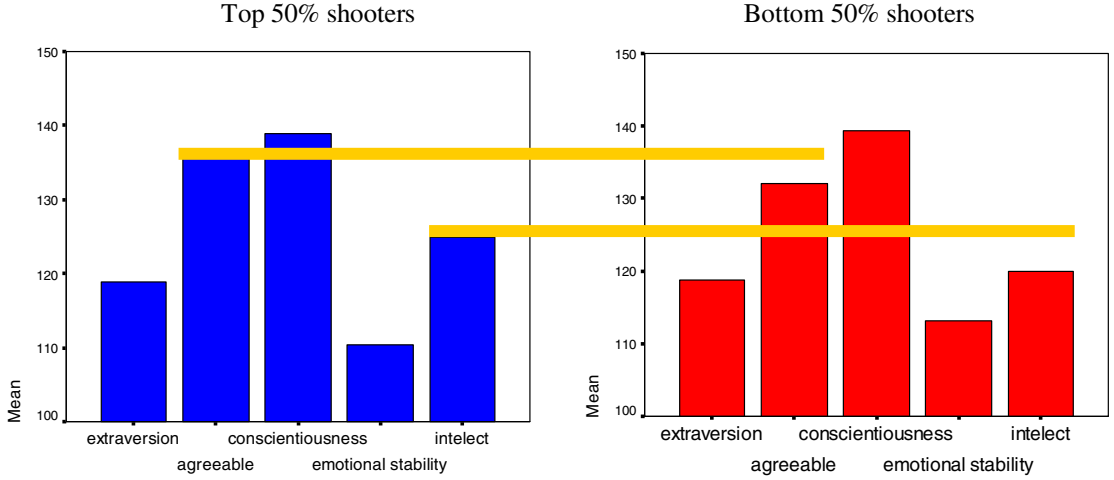
Task (I) – (J)	Mean (SD)	t	Sig.	Cohen-d	Correlation
A – B	9 (19)	4.1 (70)	.001	.44	**.58***
A – C	16 (20)	6.9 (70)	.001	.87	**.47***
A – D	11 (25)	3.5 (65)	.001	.64	.10
B – C	7 (22)	2.8 (70)	.007	.38	**.45***
B – D	2 (24)	0.6(65)	.574	.10	.21
C – D	-6(24)	-1.9 (65)	.061	.34	.15

Personality and Individual differences

We divided the shooters data into two groups where the criterion was overall performance above or below the median (median was average hit rate of 65%).

Marginally significant differences in between groups were found for *Agreeableness* (M=140.0, SD=15.0 and M=129.9, SD=19.3, accordingly) *t(25)=2.01, p<.055, Cohen's d=0.59* and for *Intellect* (M=129.9, SD=17.6 and M=119.0, SD=23.9, accordingly) *t(25)=1.95, p<.065, Cohen's d=.51* as shown in Figure 1.

Figure 1.
The Big-Five profile for shooting performance (median split for performance).



Time Perception

Time estimations were made under the prospective paradigm, where the officers knew in advance that they will be asked to judge the duration of the task. Prospective time estimations were made for Task C and Task D separately, these two tasks were also counterbalanced between group 1 (those who performed Task C before they performed Task D) and group 2 (those who performed Task D before they performed Task C).

. The average actual time for Task C was 59 seconds (SD=13), and for Task D was 57 seconds (SD=19). The average estimated time for Task C was 64 seconds (SD=44), and for Task D was 66 (SD=51). Actual time and estimated time scores were used to compute two time estimation measures. These were the duration judgment ratio (DJR; Block, Zakay, & Hancock, 1999), which is the ratio between the estimated time and the actual time in minutes and seconds (in percentages), and the absolute error which represents the distance between the estimated time and the actual time (also in percentages). Descriptive statistics for both measures are given in Table 2 and Table 3. From the descriptive statistics it can be seen that approximately 60% of the officers underestimated the duration of the task (thought that the task was shorter than it actually was), however, the variability between officers was large and the absolute error measure (Table 3) indicates that on average the time estimations were quite distant form the actual time (over 40%). Underestimation of time is quite expected considering that the time estimation was not the primary task (Zakay, 1998). The correlation between the DJR of the two tasks was significant (r = .456, *p*<.001), and respectively, the correlation between the Absolute error of both task was also significant (r =.288, *p*<.05). Thus, officers were relatively consistent in the estimation of time.

Table 2
Descriptive statistics of the DJR for Task A and Task B

Task	N	Mean (std. dev)	Min	Max	Median	Percentage of overestimations
Task C	71	109 (68)	21	450	93	43.5
Task D	69	116 (70)	13	380	94	40.8

Table 3
Descriptive statistics of the absolute error of time estimation for Task A and Task B

Task	N	Mean (std. dev)	Min	Max	Median
Task C	71	44 (52)	2.59	351	44
Task D	69	49 (53)	.67	280.17	34

Perceived Stress

Participant's responses on the pre- and post-DSSQ were used to calculate outcome scores (OS) on eleven scales using the formula (post-score–pre-score/SD), where the standard deviation used was from a large normative group reported by Matthews et al. (1999), except the two motivation scales, for which normative values were not available. For these scales, sample standard deviations were used (Matthews, Campbell, & Falconer, 2001). Significant effects were observed for Energetic Arousal $t(64)=3.09$, $p<.005$, $MSE=.10$, reflecting differences in Task Engagement. A significant effect was also observed for the Tense Arousal $t(64)=5.31$, $p<.001$, $MSE=.14$, Hedonic Tone $t(64)=-2.47$, $p<.05$, $MSE=.13$ and Control & Confidence $T(64)=-2.37$, $p<.05$, $MSE=.15$, reflecting differences in Distress. For Hedonic Tone and Control & Confidence we found moderate correlation with the overall performance scores (collapsed over the four shooting tasks) (0.26, 0.41, accordingly) implying that those who performed better felt less distressed. There were no significant differences for Worry or for Motivation.

DISCUSSION

Time estimates, especially of short durations involve cognitive processes that are sensitive to contextual conditions. The variability in time estimations, and the relative mixed literature on the topic of time estimation provided for no discernable outcome in the results for this part of the experiment. The unexpected findings for agreeableness and intellect in relationship to performing better on gun range task using the adjective checklist (Goldberg, 1992) provides an avenue for future research questions. A big 5 metaanlysis conducted by Salgado (1998) using army studies found very little information about agreeablness and intellect. Agreeableness which uses adjectives such as kindness, cooperative, considerate, pleasant. Possibly, the implications for understanding the effects of participating in a cooperative task and how that relates to the outcome of shooting performance can be predicted by agreeableness. Intellect uses adjectives such as intellectual, creative, complex, bright. Intellect may possibly be related to the nature and complexity of the task. Our future studies will attempt to explore and further reexamine these findings.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. P.A. Hancock, *Principal Investigator*. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant. The assistance of the Psychology Department at University of Central Florida is also gratefully acknowledged.

REFERENCES

Block, R. A., Zakay, D., & Hancock, P. A. (1999). Developmental changes in human duration judgments: A meta-analytic review. *Developmental Review,* **19***,* 183-211.

Easterbrook, J.A., (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review,* **66**, 183-201.

Fair, D.G. (1984). Temporal distortion. Unpublished manuscript, Department of Safety Science, University of Southern California. Los Angeles, CA.

Federal Bureau of Investigation National Press Office. (2003). *Fact Sheet for Law Enforcement Officers Killed and Assaulted, 2002.* Washington, DC: Federal Bureau of Investigation National Press Office.

Goldberg, E.R. (1992). The Developers for the Markers of the Big-Five Factor Structure. Psychological Assesment, vol. 4, No. 1, 26-42.

Hancock, P.A. and Desmond, P.A. (2001). *Stress, Workload, and Fatigue.* Mahwah: Lawrence Erlbaum Associates.

Hancock, P.A. and Parasuraman, R. (2001). Adaptive Control of Mental Workload. In Hancock, P.A. and Desmond, P.A.(Eds), *Stress, Workload, and Fatigue* (1st ed., pp. 305-320). Mahwah, NJ: Lawrence Erlbaum Associates.

Hancock, P.A. and Warm, J.S. (1989). A dynamic model of stress and sustained attention. Human Factors, **31**, 519-537.

Hancock, P.A., & Weaver, J.L. (2004). Temporal distortions under extreme stress. *Theoretical Issues in Ergonomic Science*. in press.

Masters, R.S.W. (1992). Knowledge, Knerves, and Know-how: The role of explicit versus implicit knowledge in the breakdown on a complex motor skill under pressure. *British Journal of Psychology*, ***83***, 343-358.

Salagado, J. F. (1998). Big Five Personality Dimensions and Job Performance in Army and Civil Occupations: A European Perspective, Human Performance, 11 (2/3), 271-288.

Toch, H. (2002). *Stress in Policing.* Washington, DC: American Psychological Association.

Zakay, D. (1998). Attention allocation policy influences prospective timing, *Psychonomic Bulletin & Review*, **5(1)**, 114-118.

WORKLOAD AND PERFORMANCE: A FIELD EVALUATION IN A POLICE SHOOTING RANGE

Adams Greenwood-Ericksen, Tal Oron-Gilad, James L. Szalma, Shawn Stafford, and Peter A. Hancock
University of Central Florida

One vital responsibility of all law enforcement officers lies in their obligation to defend themselves and those around them in the event of a violent threat. The best tool available to them to perform this duty is their sidearm and their skill in using it. To evaluate the efficiency of this tool use, seventy-one police officers participated in a field study examining the relationship between workload and shooting performance. Use of linear regression analysis produced evidence linking higher reported workload on the mental subscale of the NASA-TLX while completing a shooting task to poorer shooting performance when controlling for marksmanship skill. This finding is discussed in the context of previous work in the areas of workload and attentional allocation.

INTRODUCTION

Law Enforcement Officers have a difficult and dangerous job. Frequently, they operate in hostile environments and spend considerable amounts of time confronting various levels of threat. Moreover, the possibility exists that many of these situations can quickly escalate to deadly violence. Law Enforcement Officers (LEOs) face a high likelihood of violent death or injury at the hands of criminals. In the United States in 2002, 56 law enforcement officers were killed in the line of duty. In 2001, the number was 70. In 2002, 58,066 LEOs were assaulted, and 16,494 of these sustained serious injury (FBI, 2003).

Most LEOs are taught that their primary line of defense against violent assault is their firearm and the skill and judgment with which they use it. It is vitally important that LEOs be able to effectively and decisively deal with physical threats from aggressors, both to themselves and to other members of the public. Therefore officers are required to complete routine firearms assessment and training programs such as the one described in this study.

Also, officers can experience high workload demands in the line of duty, especially in emergency and threatening situations (Toch, 2002), and it is important to understand the interactions between an officer's perceived workload and their performance while using firearms. Hancock and Warm (1989) have argued that increased workload results in decreased adaptability, often resulting in poorer performance. Additionally, workload is often perceived as being greater when individuals perform tasks which are difficult or new to them. Workload would therefore be expected to decrease as officers gained more experience with their sidearms (Hancock & Desmond, 2001). Thus, officers who have more experience and greater proficiency with firearms are likely to experience less perceived workload than officers with less experience and proficiency. This leads to the inference that workload should be correlated with shooting performance such that an increase in workload leads to a decrease in shooting performance.

Due to the constrained nature of laboratory studies, field evaluations are of great importance to the study of workload since they provide a realistic setting for research and observation. The purpose of this study was to examine the perceived workload and performance in an actual police training exercise. In particular we

foresee two major advantages to this study: a) the participants were police officers, familiar with the task and with the general protocol of police shooting exercises, and b) officers' performance in this training session has a significant impact on their professional evaluation. Therefore, the purpose of the present study was to examine the hypothesized link between workload and shooting performance competence of fully-qualified police officers in a real-world and ecologically-valid evaluative setting.

EXPERIMENTAL METHOD

Experimental Participants

Participants were 71 police officers (10 women and 61 men) from a police department in the southeastern United States. The average age was 37 years (range 22-56 years). Police officers were required to participate in the evening night shooting exercise as part of their continued training. However, they volunteered to participate in the research while performing the shooting exercise.

Firearms

Each officer used a SIG (Schweizerische Industrie Gesellschaft) Sauer p226 9mm handgun. All officers used the same standard issue duty belt with identical magazine pouches. The officers carried 2 magazines with 8 and 7 rounds, respectively. Each officer also carried a standard department issue Streamlight Stinger flashlight 7.4 inches long and weighing 10 oz. As a general rule, unless they were instructed to be carried in the hand, flashlights were stored in a holder behind the radio pouch on the opposite side of that on which the officer's weapon was holstered. All officers were required to wear hearing protection (external muffs, or internal plugs), body armor, and clear eye protection.

Task Types

The task discussed in the present work was one part of a larger study involving three other shooting tasks. Of the four tasks, officers showed the worst shooting accuracy and reported the highest workload when completing the "barrel" shooting task described in this study (see description below). For this reason, this task was chosen to examine the effects of workload on performance.

'Barrel' Shooting Task Design

The 'Barrel' shooting task was either the third or the fourth task during the session (task 3 and task 4 were counterbalanced). This shooting task consisted of five human silhouette targets, distributed in a row over approximately 90 feet. Behind the targets was a 20 foot berm designed to safely absorb expended rounds. Officers began the task by facing away from the first of the five targets with a loaded weapon. On the instructors command, they engaged five silhouette targets at different ranges with a specific number of rounds per target (see Table 1). After engaging each target, officers ran to the next shooting location to engage the next target. On targets two and four, participants took cover behind plastic barrels and illuminated the target with their flashlight prior to engaging it. All firearms handling and operation was performed by police officers under the supervision of the department's training sergeant.

Simple Target Shooting Task

At the very beginning of the shooting segment of the study, each officer completed a round of familiarization fire and then proceeded to execute a simple "draw and shoot" firing task at known distances from the targets. This task is regularly performed by police officers as part of their training, and was chosen to serve as a baseline indicator of officer shooting ability

Performance Assessment

Officers' shooting performance was assessed by scoring each shot on target as a "hit," for a shot placement which was expected to kill or incapacitate an opponent, or a "miss," which was a shot placement which missed the silhouette target or was inconsistent with a mortal or incapacitating wound. Hit counts were made after each officer performed the task, and were recorded on a per-target and per-total basis.

Table 1
Number of shots and possible score per task element for the 'Barrel' task

Task Element	Range to target (yards)	Number of shots taken	Maximum possible score
target 1	25	3	3
target 2	25	4	4
target 3	7	3	3
target 4	7	3	3
target 5	5	2	2
overall task	N/A	15	15

Procedure

Six to twelve officers were assigned to each shooting exercise at a time, but each officer performed the task separately (in a sequential order). The training session took place in the dark at an outdoor police shooting range, and the total session time varied from approximately 60 minutes to 120 minutes depending on the number of officers in the session. All officers were given the same instructions for procedural and safety reasons. Additionally, the officers were told that they would be asked by the experimenter to provide an estimate of the task length in minutes and seconds upon completion of the task. Each officer filled out the NASA-TLX questionnaire after completion of the task.

RESULTS

All statistical analyses were performed using SPSS for windows, version 11.5. Unless otherwise stated, an alpha level of .05 was used for all analyses. Tables 2 and 3 below show group statistics for TLX and 'Barrel' shooting task performance.

Table 2
NASA-TLX Global and Subscale Descriptives for the 'Barrel' task

TLX element	N	Mean	Std. Dev.
Mental subscale	70	67.96	20.87
Physical subscale	70	61.89	23.40
Temporal subscale	70	65.30	22.88
Performance subscale	70	44.11	22.84
Effort subscale	70	63.90	21.29
Frustration subscale	68	44.46	28.49
Mean Score	70	57.73	12.96

Table 3
Overall shooting performance and Performance by target for the 'Barrel' task

Task Element	N	Mean Score	Std. Dev.	Maximum possible score
target 1	44	1.5	1.00	3
target 2	44	2.02	1.29	4
target 3	44	2.07	0.95	3
target 4	44	2.07	1.11	3
target 5	44	1.14	0.73	2
overall task	71	8.75	3.05	15

Note. Individual target data was only available for participants 28 through 74. Three participants dropped out of the study prior to completion

A hierarchical linear regression was performed on shooting performance with total targets hits as the dependent variable and with shooting skill (assessed in the baseline shooting task) entering first, followed by workload

(overall score on the NASA-TLX). After shooting ability was entered, the initial model was significant, F (1, 68) = 19.433, $p < .0005$, and predicted about 22.2 % of the overall variance (R^2) in shooting performance on the barrel task. The analysis indicated the following predictive equation for the model:
Y = 36.89– 2.61x.

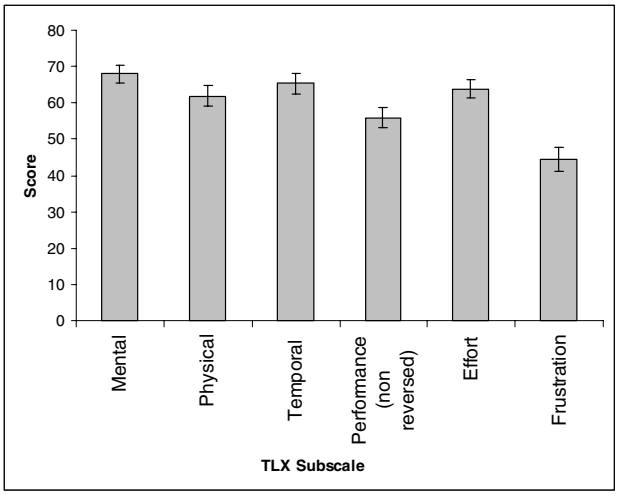
When the overall workload score was entered into this model, it increased the predictive power by 2.7%. However, the change failed to reach significance, indicating that overall workload score and shooting ability together did not account for significantly more variance in shooting performance than shooting skill alone.

In order to test the a priori hypothesis regarding the relationship between each of the participants' scores on the six individual TLX workload subscales and shooting performance, a series of hierarchical linear regressions were performed on shooting performance with shooting skill entering first in each case followed by one of the individual subscales. In each case, the overall model was significant, but only the model in which *mental* workload was entered showed a significant improvement in predictive power (R^2) over the model with shooting skill alone, F (2, 67) = 13.525, $p < .0005$. The model accounted for 28.8% of the overall variance in shooting performance, 6.5% more than the variance accounted for by shooting skill alone. The model yielded the following equation: Y = 39.57 – 2.51x.

DISCUSSION

The results indicated that the *mental* demand subscale of the NASA-TLX is significantly correlated to shooting performance above and beyond the variance due to shooting skill. This finding is open to a number of plausible interpretations. It seems likely that officers who are less skilled in handling firearms are most likely to experience increased compensatory workload while using them. Increases in workload may be expected under these conditions because these officers do not handle or discharge their weapon on a routine basis. Such increases resulting from task demand have been associated with poorer performance (Parasuraman & Hancock, 2001). Thus, this explanation would posit that officers who are less skilled devote more attentional resources to compensate for their lack of skill, but still perform at a lower level than more skilled officers. However, other plausible explanations exist for this phenomenon.

Figure 1
NASA-TLX Subscale Scores and Std. Errors for the 'Barrel' Shooting task



Alternatively, it is possible that devoting additional mental resources to shooting performance might have paradoxically negative effects on performance, as officers "over-think" the task and "choke" under pressure (see Masters, 1992). This explanation is consistent with the view that implicit skills gained through practice break down under pressure as individuals devote additional attentional resources in an attempt to improve performance. In this case, officers attempt to substitute explicit knowledge of how to use the weapon gained during their formal training for implicit knowledge gained through practice. This explanation suggests that poorer performing officers might not be less skilled, but are simply less resilient to pressure to perform. This finding casts doubt on the common training procedure of testing officers in situations in which they are

likely to feel pressure to perform, as it is possible that some officers might perform more poorly during assessments than they would during actual firefights as a result of their explicit focus on the mechanics of shooting during these procedures.

The finding that only one subscale of the TLX is significantly correlated to performance on the shooting task presents an interesting quandary. Given the physical nature of the task, involving running and taking cover while engaging targets, we might expect that the *effort* and *physical* demand subscales would show an effect. The finding that *mental* demand accounted for such a large amount of variance in shooting performance indicates that that those officers who expended the most mental effort on the task did the poorest. This may be an example of the phenomenon of performance-workload dissociation which has been discussed at length by Hancock (1996). Lower performing officers may have felt that they had to exert more effort, supporting the hypothesis that these officers were allocating more resources to the task, resulting in paradoxically poorer performance due to choking.

Unfortunately, because of the necessarily correlational design of this study, it is impossible to determine whether the workload is the cause or the symptom of poor shooting performance. It is possible that future studies may find a way to determine the causal relationship between these two constructs, and this represents an area which should be addressed in future empirical work.

ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621, Peter A. Hancock, Principal Investigator.

REFERENCES

Federal Bureau of Investigation National Press Office. (2003). *Fact Sheet for Law Enforcement Officers Killed and Assaulted, 2002.* Washington, DC: Federal Bureau of Investigation National Press Office.

Hancock, P.A. (1996). Effects of control order, augmented feedback, input device and practive on tracking performance and perceived workload, *Ergonomics*, **39**(9), 1146-1162

Hancock, P.A. and Warm, J.S. (1989). A dynamic model of stress and sustained attention. Human Factors, **31**, 519-537.

Hancock, P.A. and Desmond, P.A. (2001). *Stress, Workload, and Fatigue.* Mahwah: Lawrence Erlbaum Associates.

Hancock, P.A. and Parasuraman, R. (2001). Adaptive Control of Mental Workload. In Hancock, P.A. and Desmond, P.A.(Eds), *Stress, Workload, and Fatigue* (1st ed., pp 305-320). Mahwah, NJ: Lawrence Erlbaum Associates.

Masters, R.S.W. (1992). Knowledge, Knerves, and Know-how: The role of explicit versus implicit knowledge in the breakdown on a complex motor skill under pressure. *British Journal of Psychology*, *83*, 343-358.

Toch, H. (2002). *Stress in Policing.* Washington, DC: American Psychological Association.

PERFORMANCE OPERATING CHARACTERISTICS FOR SPATIAL AND TEMPORAL DISCRIMINATIONS: COMMON OR SEPARATE CAPACITIES?

J. E. Thropp, J. L. Szalma, and P. A. Hancock
MURI-OPUS Research Group
MIT^2 Laboratories
University of Central Florida
Orlando, FL

Resource-sharing between spatial and temporal processing was investigated using dual task methodology to construct performance operating characteristics. Spatial tasks involved discrimination of line length and temporal tasks involved discrimination of duration. Preliminary results suggest that a tradeoff occurs within the easier task of the two difficulty levels, but that a simple tradeoff is not observed at the more difficult level. This pattern of results suggests that the relationship between spatial and temporal perception varies according to difficulty level. That is, the degree to which the processes share common capacities varies as a function of the metrical characteristics of stimuli to be processed.

INTRODUCTION

Hancock and Warm (1989) have argued that performance tasks which face the operators are the proximal sources of stress. They articulated two general dimensions along which tasks vary, these being information structure and information rate. Information structure refers to task elements that are often represented spatially, while information rate predominately refers to the temporal characteristics of the task. According to Hancock and Warm (1989), one way operators cope with these demands is by narrowing their attention to specific environmental cues (Easterbrook, 1959). Hancock and Weaver (in press) argued that the narrowing of attention, which has been associated with distortions of spatial perception (Easterbrook, 1959), also produces distortion along the temporal dimension. Thus, distortion of space and time may share a common mechanism, and this may result from a common resource capacity for processing spatial and temporal information. This study examined this possibility by constructing performance operating characteristics (POCs) using a dual task methodology (Navon & Gopher, 1979). If these two processes draw upon common resource capacities, a tradeoff in performance should be observed as attention is allocated to either task.

EXPERIMENTAL METHOD

Experimental Participants

Six students, three females and three males, enrolled at the University of Central Florida, served as participants in order to receive a financial reimbursement. They ranged in age from 18 to 24 years, with a mean age of 20. All participants completed a set of ten experimental conditions consisting of various attentional allocation strategies and difficulty levels. A total of four days of participation were required, all of which were completed within a span of one week.

Experimental Stimuli

All participants engaged in spatial and temporal line discrimination tasks. In the spatial discrimination task, participants differentiated between vertical lines which were either short (32 mm) or tall (35 mm in the easy version and 33 mm in the difficult version). In the temporal discrimination task, they were required to differentiate between lines presented at either a short (300 ms) or long (450 ms in the easy version and 360 ms in the difficult version) duration. All lines were presented on a computer monitor.

There were a total of ten experimental conditions which reflected five attentional allocation strategies for two difficulty levels. Participants were asked to allocate a relative proportion of their attention to each dimension. For each level of difficulty, there were two baseline conditions in which only one of the dimensions was assessed: a 100% Spatial/0% Temporal Baseline condition and a 100% Temporal/0% Spatial Baseline condition. Thus, in these baseline conditions, the participants were instructed to allocate 100% of their attention to one of the dimensions while completely disregarding the other dimension. There were also three dual task conditions for each difficulty level: 90% Spatial/10% Temporal, 90% Temporal/10% Spatial, and 50%Spatial/50% Temporal. Each condition consisted of 480 trials; thus 480 lines were presented in each condition. The large number of trials was necessary to obtain stable signal detection theory

estimates (Green & Swets, 1966). Data for each participant was collected in four two-hour sessions over four days.

Pilot testing was completed in order to achieve psychophysical equivalency both the spatial and temporal tasks in terms of difficulty. Participants completed the two computerized discrimination tasks involving a two-alternative forced choice procedure. These two tasks were either spatial or temporal in nature. Lines of different heights and presentation times were presented in pairs, and participants indicated which line was either taller or presented for a longer duration.

Procedure

Participants began with a practice session consisting of two-alternative forced choice tasks, which were similar to the tasks used in the pilot testing. Upon completing practice sessions, participants completed the line length and duration discrimination conditions.

The first day of testing consisted of four conditions: an easy and difficult version of both the spatial and temporal baseline conditions. In the 100% Spatial/0% Temporal Baseline condition, participants were instructed to ignore changes in duration and instead devote all of their attention to monitoring the height of the lines. After each line was presented, a screen appeared asking participants to indicate the height of each line. Participants responded using a four-point confidence rating whereby they pressed the 1 key on the keyboard if the line was *definitely short* in height, the 2 key if the line was *possibly short* in height, the 3 key if the line was *possibly tall* in height, the 4 key if the line was *definitely tall* in height. Thus, participants were asked to indicate the confidence with which they felt each line was short or tall in height.

The 100% Temporal/0% Spatial Baseline condition followed the same format as the Spatial Baseline, except participants were instructed to ignore changes in the heights of the lines and instead devote all of their attention to monitoring the presentation times. A similar four-point confidence rating scale was used. All four baseline conditions were administered on the first day. The administration orders of all four conditions were counterbalanced between participants.

The second, third, and fourth days consisted of the easy and difficult versions of one of the dual-task attentional allocation strategies: 90% Spatial/10% Temporal, 90% Temporal/10% Spatial, and 50%Spatial/50% Temporal. The administration orders of all three conditions were counterbalanced between participants. In the dual task conditions, a line was presented and two screens appeared, one after the other and in random order, asking the participant for confidence ratings in either time or height. Thus, each line was rated for both time and height by the participants on each trial. There was a spatial and temporal rating for each presentation of the line stimulus.

RESULTS

Sensitivity scores for each task were computed separately for each participant from the performance operating characteristic (POC) curves derived from the ratings data using maximum likelihood procedures described by MacMillan and Creelman (1991). This technique permits statistical analysis of differences between conditions based on a large number of trials per condition.

Figure 1 depicts the POCs of average perceptual sensitivity in the spatial task as a function of average perceptual sensitivity in the temporal task. Figures 2 through 7 depict perceptual sensitivity in the spatial task as a function of perceptual sensitivity in the temporal task for each of the six participants.

Spatial and temporal baseline conditions were assessed for each participant. In the easy condition, the baselines were significantly different for participant 2, $p<.0001$, participant 4, $p<.005$, participant 5, $p<.0001$, and participant 6, $p<.01$. Baseline performance in the easy condition were statistically equivalent for participants 1 and 3, ($p<.05$). In the difficult condition, baseline scores were significantly different for participant 2, $p<.0001$ and participant 4, $p<.05$. Baseline scores in the difficult condition were statistically equivalent for participants 1, 3, 5, and 6 ($p>.05$).

Results suggest that a tradeoff occurs within the easier task of the two difficulty levels, but that a simple tradeoff is not observed at the more difficult level. This pattern of results suggests that the relationship between spatial and temporal perception varies according to difficulty level. That is, the degree to which the processes share common capacities may vary as a function of the range of durations and/or heights to be discriminated. Indeed, prior research in the intensity of auditory signals has indicated that the range of stimulus values employed can influence detection performance (Braida & Durlach, 1972).

In general, many of the subjects exhibited a degree of tradeoff in performance in the dual task conditions. However, as shown by some of the POC graphs, performance on one dimension was often facilitated by the concurrent performance of the other dimension. Yet, this trend was evident in only a few of the participants; this indicates the role of individual differences in time-sharing, dual task performance, and spatiotemporal ability.

Figure 1. *Performance operating characteristics (POC) of average perceptual sensitivity in the spatial task as a function of average perceptual sensitivity in the temporal task*

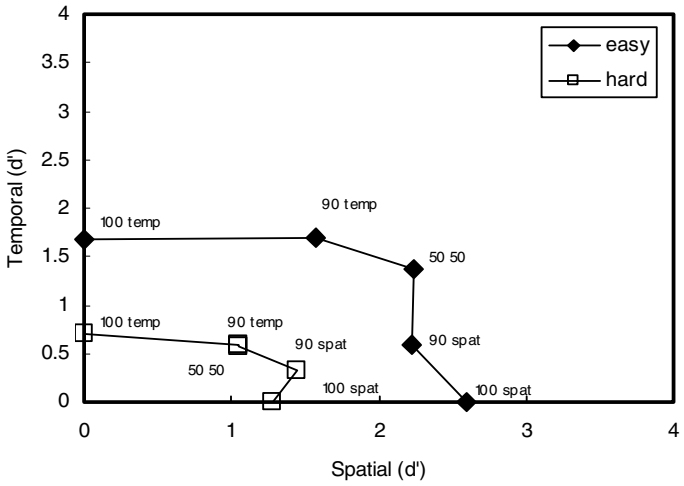


Figure 2. *POC for Participant 1*

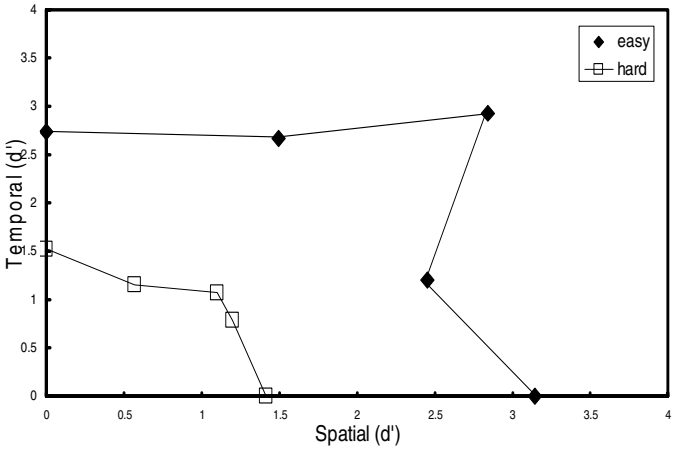


Figure 3. *POC for Participant 2*

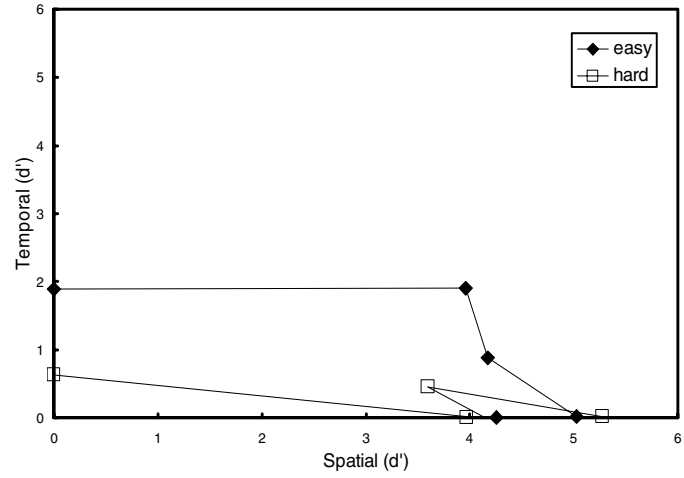


Figure 4. *POC for Participant 3*

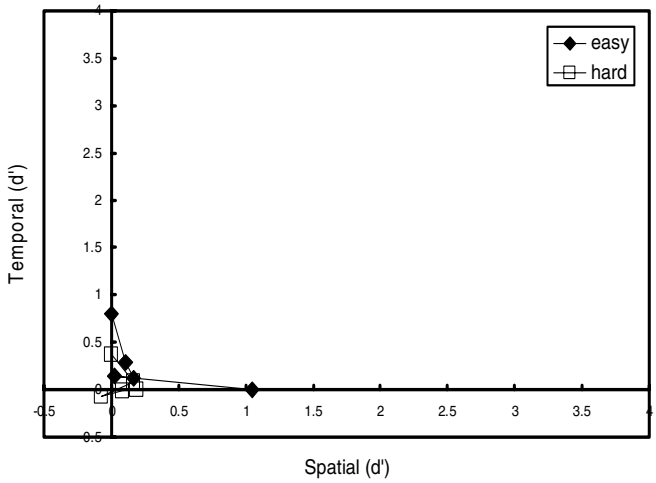


Figure 5. *POC for Participant 4*

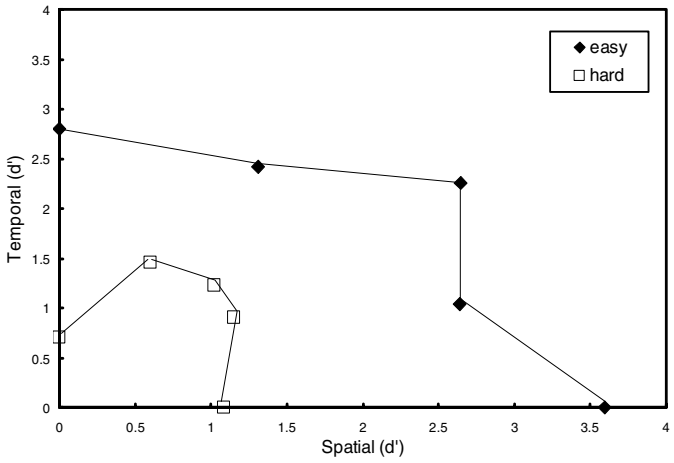


Figure 6. *POC for Participant 5*

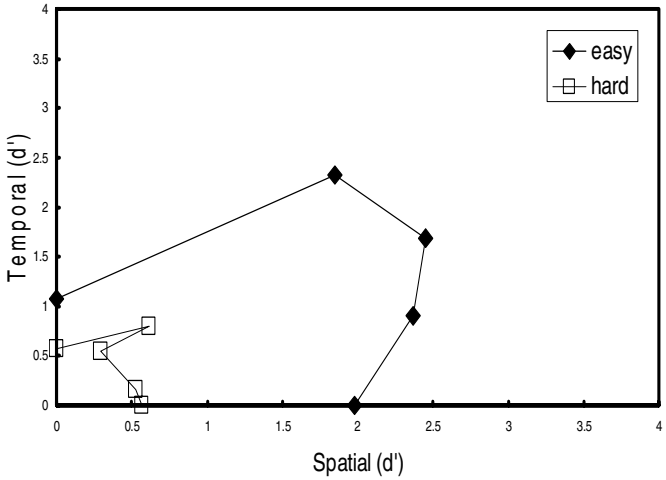
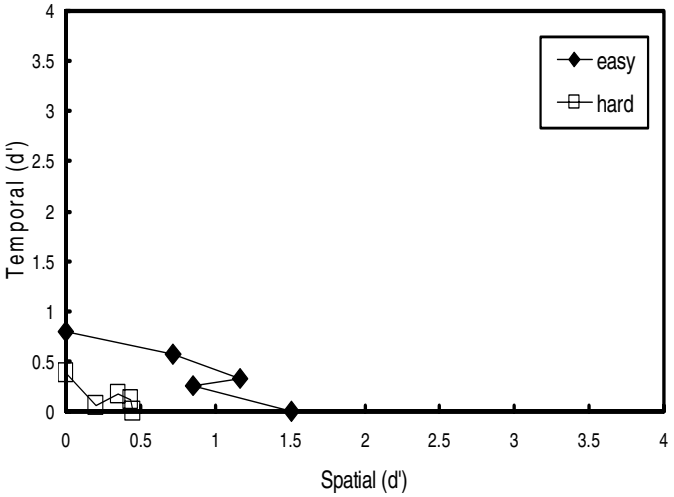


Figure 7. *POC for Participant 6*



DISCUSSION

Inspection of the POC curve based upon data averaged across the participants suggests a partial tradeoff, in which the spatial task is more vulnerable to the attention allocation manipulation than the temporal task. That is, sacrificing attention to the spatial task in favor of the temporal task does not improve beyond the 50/50 condition, evident from the small change in performance on the temporal task from the baseline to the 90% Temporal to 50% Temporal conditions (see Figure 1). Following Navon and Gopher (1979), the shape of the POC curve (easy condition) based on averaged data indicates that the spatial task draws on resource capacities shared with the temporal task, but the latter task also draws on a separate resource pool that is independent of the spatial dimension.

Note that these effects do not imply that the temporal dimension utilizes processing capacities that are independent of other perceptual domains. Although there is some non-overlap in resource utilization between the spatial and the temporal dimensions, prior research has indicated that processing of temporal information does share resource capacities with processing luminance information (see Grondin & Macar, 1992).

The above interpretation of the results of this study, based on averaged data, masks the large differences between participants in performance and in the shape of the POC. Inspection of the data for each individual participant indicates that the averaged data masks vast individual differences in overall performance and in ability to effectively allocate attention. In one case (participant 2) performance on the spatial task far exceeded that on the temporal task, while another participant (participant 3) was unable to perform either task very well. For the difficult condition, a tradeoff between tasks was observed for only one participant (#1), while for the easy condition tradeoffs were observed for three of the observers (participants 1, 4, and 6). Taken at face value, these results suggest that the degree of resource overlap between the spatial and temporal discrimination tasks depends on both the difficulty level of the task and the individual engaged in the task. There is also a potential for a gender effect, since two of the observers whose data showed the most consistent data were women (participants 1 and 4).

It is also important to note that baseline performance for the two tasks (i.e., performance of the spatial task and the temporal task under single task conditions) were equivalent only for the difficult conditions for participants 1 3, 5, and 6, and in the case of the latter three participants the baseline performances were not significantly different from performance in the dual task conditions. Thus, in the easy conditions (participants 2, 4, 5, and 6) and in the difficult conditions for participants 2 and 4, the baseline sensitivity scores were significantly different in spite of the fact that the discriminations were psychophysically equated under alerted conditions. Thus, the equivalency of discriminations may also depend upon characteristics of individual observers and how they respond to different levels of task difficulty.

A central question therefore is what observer characteristics are most important in determining dual task performance and the subsequent shape of POCs. At least four individual differences variables may have influenced the results of the current experiment: 1) Time sharing ability, 2) Ability to consciously control one's allocation of attention (c.f., Derryberry & Reed, 2001), 3) Ability to do spatial discrimination, and 4) Ability to do temporal discrimination. It is most likely that individual differences relating to task-specific characteristics drive differences among participants, since prior research has indicated that there is limited evidence for a general time sharing ability factor (e.g., Wickens, Mountford, & Scheriner, 1981). Efforts are currently underway to identify those specific individual differences that relate to processing spatial and visual information of the kind examined in this study. Based on the findings of this experiment, an individual differences approach to POC studies seems crucial to understanding the effects of manipulation of attention allocation and task difficulty.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program, P.A. Hancock, *Principal Investig*ator, administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

REFERENCES

Braida, L. D. & Durlach, N. I. (1972). Intensity Perception II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America, 51,* 483-502.

Derryberry, D., & Reed, M.A. (2001). A multidisciplinary perspective on attentional control. In C.L. Folk and B.S. Gibson (Eds.), Attraction, distraction, and action. Amsterdam: Elsevier.

Easterbrook, J. A. (1959). The effect of motion on cue utilization and the organization of behavior. *Psychological Review, 66,* 183-201.

Green, D. & Swets, J. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Grondin, S. & Macar, F. (1992). Dividing attention between temporal and nontemporal tasks: A performance operating characteristic POC analysis. In F. Macar, V. Pouthas, and W. Friedman (Eds.) Time, Action, and Cognition: Towards bridging the gap. Dordrecht, Netherlands: Kluwer Academic.

Hancock. P. A. & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31(5),* 519-537.

Hancock, P. A. & Weaver, J. L. (in press). On distortion under time stress. *Theoretical Issues in Ergonomic Science.*

MacMillan, N.A., & Creelman, C.D. (1991). *Detection Theory: A User's Guide.* Cambridge: Cambridge University Press.

Navon, D. & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86,* 214-255.

Wickens, C. D., Mountford, S. J., & Schreiner, W. (1981). Multiple resources, task-hemispheric integrity, and individual differences in time-sharing. *Human Factors, 23(2),* 211-229.

THE EFFECTS OF WARNING PRESENTATION AND RETENTION UNDER VARYING LEVELS OF STRESS

Jessica Helmick-Rich, Kelly A. Burke, Tal Oron-Gilad, and Peter A. Hancock
University of Central Florida

The modality in which the information is presented differentially influences behavioral response, especially when tasks are either learned or subsequently performed in stressful circumstances. The current work examines the effects of cross modality of warning presentation and retention in a dual task paradigm in a simulated environment with various task-induced stress levels. Consistent with the Hancock and Warm model of stress and attention, it was found that when task demand is relatively low, the modality of presentation is of less importance and participants are able to comply. When task demand is relatively high, the modality of presentation was critical and played a significant role in compliance behavior. Additionally, it was found that participants were significantly less likely to comply in the verbal modality across all levels of working memory demand than in either the pictorial or written modalities.

INTRODUCTION

Warnings communicate hazards and are intended to prevent accidents and injuries from occurring. Warnings are usually not presented in isolation, but are generally presented while the operator is performing a certain operational task. Warning messages should attract the operator's attention, inform the operator of the hazard, and hopefully convince the operator to comply. The format in which warnings is presented is not standardized; warnings vary in their presentation format across tasks and environments. Presentations of hazard warnings are commonly found in one of the following formats: pictorial, written, or auditory. Consequently, the literature pertaining to modality is unclear as to which types of presentation is most memorable, salient, or most effective resulting in a higher rate of compliance behavior under stressful conditions. In the current study we manipulated warning presentation and warning retention. Presentation was manipulated by the modality of presentation of the warning cues (verbal, written or pictorial). Retention was manipulated by creating associations between the hazard and the warning cue through different modalities (verbal, written or pictorial). Hence this study covers not only the effect of modality on warning compliance but also the effect of working memory associations (spatial and verbal associations) on warning retention.

Historically, research in the warning modality domain concentrated on comparing verbal-written presentations or picture-word differences. Penney (1975) concluded that auditory presentation (spoken words) of information results in better retention then visual presentation (written words) in short term memory tasks. Others suggested that people can recognize and recall pictures more quickly and accurately than words (Ells & Dewar, 1979; Janda & Volk 1934; Paivo, Rogers, & Smythe, 1968; Standing, Conezio, & Haber, 1970). Several theories support the claim that pictures are processed both verbally and symbolically, therefore having an advantage over words. (Pavio, 1971; 1975; & 1978; Siegel & Dhawan, 1975). Similarly, literature in text processing found pictorials to have a processing advantage over text, thus arguing that pictures facilitate comprehension (Glenberg & Langston, 1992; Larkin & Simon, 1987). Some argue that pictorials are "instant reminders" of the hazard (Peters, 1984), more recognizable than words, (Paivio, Rogers, & Smythe, 1968; Standing, Conezio, & Haber, 1970), and they convey information more rapidly and effectively than verbal messages (Dorris & Perswell, 1978). Additionally, OSHA and NIOSH guidelines suggest adding a pictorial to the warning label. Contrary to the picture superiority findings, others (Robinett & Hughes, 1984; Smith & Magee, 1980) found that icons are sometimes complex and cannot convey the appropriate information without a written message accompanying them.

Our current study examines the effects of cross modality of warning presentation and retention in a dual task paradigm in a simulated environment. In general, attentional resources are utilized better when divided across modalities (auditory and visual stimuli, for example) rather than displayed via two auditory or two visual channels, (Wickens, 1992; Wickens, Sandry, & Vidulich, 1983). Similarly, several models of working memory are associated with different modalities or different types of representations and are consistent with the notion of multiple resource pools (Daneman & Tardif, 1987; Martin, 1993; Monsell, 1984; Shah & Miyake, 1996). Since the operational task in this study is predominantly a visual and spatial task, one can hypothesize that participants would have a significantly higher rate of compliance behavior on the warning-compliance task when the warning was presented in verbal compared to written and pictorial modality because the warning-compliance task would have less interference with the operational task.

According to the Hancock and Warm model (1989), performance will be affected by stress when it increases to the point that it is outside of the comfort zone. Under low stress conditions participants are more likely to comply than under high stress conditions (Magurno & Wogalter, 1994). In the current study, operational stress level was manipulated by the task demands of the warning-compliance task (the number of

association cues that the operator had to retrieve). In particular, it was of interest to determine the amount of task demand which still provides a high level of compliance. Therefore, it was hypothesized that the comfort zone would be maintained when a participant was presented with a maximum of four stimuli in any modality. Furthermore, when participants were presented with more than four stimuli they would fall outside of the comfort zone resulting in degraded performance.

METHOD

Participants

Sixteen participants (mean age = 21.13, *SD* = 3.68*)* were recruited on a voluntary basis from undergraduate psychology classes at the University of Central Florida and were paid approximately $37.50 for their participation (based on an hourly rate of $7.50).

Materials

Dual Task Setting. The Warning Color-Combination (WCCOM) compliance task and the shooting task were presented on two separate monitors with two keyboards and mice. The computer used for the WCCOM had two speakers which were used to present the verbal warnings. The computer used for the shooting task did not have speakers, thus no noise was emitted from during the task. The two computers were placed on a desk side-by side in order for participants to easily view both monitors.

WCCOM Compliance Task. WCCOM was one of the two tasks in the dual task paradigm. The warning-color combinations in this task consisted of one of ten warnings (boots, earmuffs, glasses, gloves, helmet, shield, suit, respirator, meter, or mask) paired with one of ten colors (red, blue, green, orange, purple, black, white, gray, brown, or yellow; recommended by the American National Standards Institute, Inc.; ANSI, 1991). The WCCOM consisted of both a storage and a processing requirement. The storage requirement consisted of learning the color associated with each warning. Warnings were presented in one of three modalities, pictorial, written, or verbal. Each combination was paired randomly and appeared only once per block.

Each warning-color combination was presented individually on the computer screen for five seconds. An example of the pictorial presentation of the WCCOM is depicted in Figure 1. In this example the warning, boots, is paired with the color black. The written warnings were presented in the same manner (see Figure 2). The verbal WCCOM was presented via speakers. For example, the participants heard "*boots...black*". Operational stress level was manipulated by the working memory demands of the WCCOM, the number of association cues that the operator had to retrieve, which increased in demand from two to four to eight.

The processing requirement of the task involved complying with the warning (responding with a keypress) while simultaneously performing the shooting task. Each trial consisted of a two-minute session of the shooting task during which the color portion of the WCCOM was presented at random times. The participant's task was to remember the correct pairing of the warning and color combinations. They responded by pressing the appropriately labeled key on a second keyboard (keys 'q' through 'p' were labeled with a pictorial representation of the warning portion of each combination) with their right hand.



Figure 1. Example of a pictorial warning-color combination (left) and the color stimulus (right) that elicits the key press response during the WCCOM portion of the dual task. In this specific example, the warning, boots, are combined with the color black.

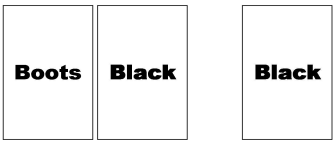


Figure 2. Example of a written warning-color combination (left) and the color stimulus (right) that elicits the key press response during the WCCOM portion of the dual task.

Shooting Task. During each trial, the participants completed a two-minute mission in Ghost Recon. During each mission, the objective was to clear a building by entering each room and killing all of the enemies. Participants navigated through the Ghost Recon environment using the arrow keys on one keyboard with their left hand and the mouse with their right hand. Performance was measured by calculating the hit percentage (number of shots which hit a target divided by the total number of shots fired).

Card Sorting Task. One complete deck of playing cards was used to administer the card sorting task. This task was used as an attempt to diminish carry over effects from the previous trial.

Procedure

An all within participants design was used which consisted of three blocks (one for each modality: verbal, written, and pictorial); each block consisted of 15 trials (5 repetitions for each memory demand level 2, 4, and 8 retrieved warnings). Presentation order of the modality conditions were counterbalanced between blocks to avoid the effect of sequencing. Additionally, the presentation order of trials within each block was randomized.

Testing occurred in two sessions (approximately 2.5 hours each) on different days during one week. During session 1 the

participants were asked to complete the informed consent, the demographic questionnaire, and the State Trait Anxiety Inventory Form Y (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). After a five-minute break, four working memory tasks (see Shah & Miyake, 1996) were administered followed by another five-minute break. The participants completed a practice session, which consisted of three trials of the shooting task, the WCCOM, and both tasks simultaneously (dual task paradigm). Following the practice sessions, participants completed the first block. Session 2 consisted of the remaining two blocks of the experiment. A five-minute break was scheduled between blocks.

During both sessions, the Rating Scale Mental Effort (RSME; Zijlstra, & Van Doorn, 1985) and the card-sorting task were administered following each trial and the NASA-Task Load Index (NASA-TLX; Hart, & Staveland, 1988) followed each block. Finally, participants were debriefed via a verbal and written statement.

RESULTS

WCCOM Task. A three-way 3 (modality) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on warning compliance behavior. The independent variables consisted of three different types of modality (verbal, pictorial, and written) in which the warning-color combinations were presented to the participant and varying levels of task demand (presentations of two, four, and eight). The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination) at each level of task demand. Analyses were performed using SPSS, for Windows, 11.0 and an alpha level of .05 was used.

There was a significant main effect of modality and task demand, Wilk's $\Lambda = .62$, $F(2, 13) = 10.8$, $p = .02$, partial $\eta^2 = .62$, and Wilk's $\Lambda = .89$, $F(2, 13) = 39.36$, $p = .005$, partial $\eta^2 = .86$, respectively, and a significant interaction between modality and task demand, Wilk's $\Lambda = .57$, $F(4, 11) = 3.65$, $p = .04$, partial $\eta^2 = .57$. However, there was not a significant main effect of trial.

A set of Fisher LSD post hoc tests for modality type show that participants were significantly more likely to comply when the information was presented in the written, ($M = 73$, $SD = 4$) and pictorial ($M = 65$, $SD = 5$) modalities than in the verbal modality ($M = 52$, $SD = 3$). Furthermore, Fisher LSD post hoc tests for working memory show that there was a significant difference in compliance behavior between the levels of two ($M = 75$, $SD = 3$) and eight ($M = 47$, $SD = 3$) and levels four ($M = 68$, $SD = 5$) and eight ($M = 47$, $SD = 3$), but not between levels two and four (see Figure 3).

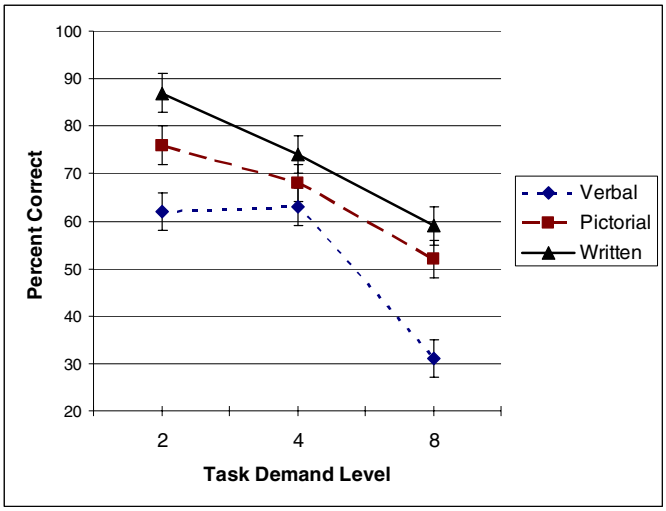


Figure 3. Relationship between response to warning cue, modality type and task demand level.

Shooting task. A three-way 3 (modality) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on percentage of hits in each trial. There was not a significant main effect of modality or task demand, Wilk's $\Lambda = .94$, $F(2, 13) = .38$, $p = .69$, partial $\eta^2 = .06$, and Wilk's $\Lambda = .87$, $F(2, 13) = .94$, $p = .42$, partial $\eta^2 = .13$, respectively, and there were no significant interactions (see Figure 4).

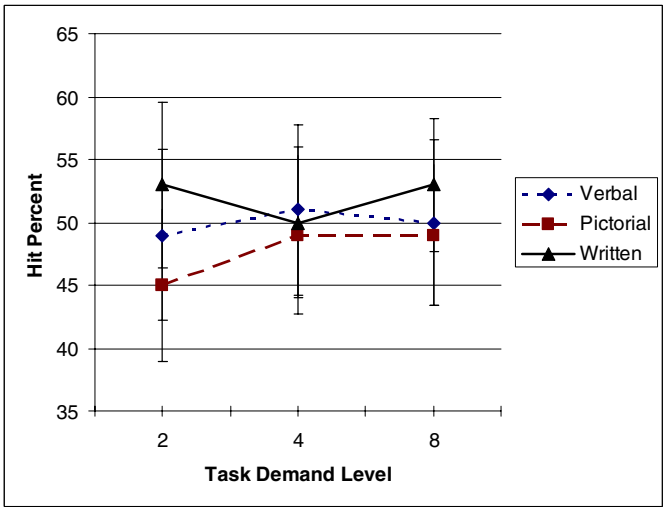


Figure 4. Relationship between hit percent, modality type and task demand level on the Ghost Recon shooting task.

DISCUSSION

Consistent with the Hancock and Warm model of stress and attention (Hancock & Warm, 1989), it was hypothesized that the comfort zone would be maintained when a participant is presented with a maximum of four stimuli in any modality. Furthermore, when participants were presented with 8 stimuli, it resulted in degraded performance. In support of this

hypothesis, it was found that when working memory demands (task stress) are relatively small (two and four warnings), the modality of presentation is of less importance and participants are able to comply. When task demands are relatively high (eight warnings) the modality of presentation was important and played a significant role in compliance behavior.

ACKNOWLEDGEMENTS

The view expressed in thesis work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-0621, Peter A. Hancock, Principal Investigator. This work was also supported in part by DARPA grant NBCH1030012, Program Manager, Dylan Schmorrow.

REFERENCES

American National Standards Institute, Inc. (1991). American national standard for environmental and facility safety signs, ANSI Z535.2-1991.

American Psychological Association (1992). Ethical principles of psychologists and code of conduct. American Psychologist, 47, 1597-1611.

Daneman, M., & Tardiff, T. (1987). Working memory and reading skill reexamined. In M Coltheart (Eds.), *Attention and Performance XII: The psychology of reading.* (pp. 491-508). Hillsdale, NJ: Erlbaum.

Dorris, A. L., & Purswell, J.L.(1978).Human factors in the design of effective product warnings. *Proceedings of the Human Factors and Ergonomics Society 33rd Annual Meeting,* 343-346.

Ells, J.G., & Dewar, R.E. (1979). Rapid comprehension of verbal and symbolic traffic sign messages. *Human Factors, 21*(2), 161-168.

Glenberg, A. M., & Langston, W. E. (1992).Comprehension of illustrated text: Pictures build mental models. *Journal of Memory and Language, 31*, 129-151.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519-537.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*, (pp. 139-183). Amsterdam: North-Holland.

Janda, H.F., & Volk, W.N. (1934). Effectiveness of various highway signs. National Academy of Sciences, National Research Council, *Highway Research Board Proceedings, 14*, 442-47.

Larkin, J.H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11,* 65-99.

Magurno, A.B., & Wolgalter, M.S. (1994). Behavioral compliance with warnings: Effects of stress and placement. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, 826-830.

Martin, R.C. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. *Memory and Cognition, 21,* 176-183.

Millisecond Software (2002). Inquisit, Version 1.32. Seattle, WA: Millisecond Software.

Monsell, S. (1984). Component of working memory underlying verbal skills: A distributed capacities' view-A tutorial review. In H. Bouma & D.G. Bouwhuis. (Eds.), *Attention and performance X* (pp. 327-350). Hillsdale, NJ: Earlbaum.

Paivo, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory and Cognition, 3*, 635-647.

Paivio, A. (1978). A dual coding approach to perception and cognition. In H.L. Pick & E. Saltzman (Eds.), Modes of perceiving and processing information (pp. 39-51). Hilsdale NJ: Erlbaum.

Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin, 82*(1), 68-84.

Robinett, F., & Hughes, A. (1984). Visual alerts to machinery hazards: A design case study. In Easterby, R., & Zwaga, H. (Eds.), *Information Design: The Design and Evaluation of Signs and Printed Material* (pp. 405-417).

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual difference approach. *Journal of Experimental Psychology: General, 125*, 4-27.

Smith, M.C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General, 109*, 373-392.

Spielberger, C.D., Gorsuch, R.L., Lushene, P., Vagg, P.R., & Jacobs, G.A. (1983). State-Trait Anxiety Inventory Form Y. Consulting Psychologists Press, Inc.

Zijlstra, F.R.H. & Van Doorn, L. (1985). The construction of a scale to measure perceived effort. Delft, The Netherlands: Department of Philosophy and Social Sciences, Delft University of Technology.

Affect as an Aspect of Workload: Why and How to Address Affect in Automation and Simulation Studies

H. C. Neil Ganey
University of Central Florida

Mark E. Koltko-Rivera
Professional Services Group
Winter Park, Florida

Lauren M. Murphy
University of Central Florida

P. A. Hancock
University of Central Florida and
Institute for Simulation and Training

Joseph Dalton
University of Central Florida

HPSAA

Affect is a much-neglected moderator of workload and performance. Research literature demonstrating affect's influence on cognitive and physical tasks is summarized. In terms of cognitive performance, affect has been shown to influence memory function, decision making, problem solving, risk assessment, and performance on other cognitively loaded tasks. In terms of physical performance, affect has been shown to influence physical perceptions and interpersonal functioning, and to moderate performance on tasks that are heavily loaded for strenuous physical activity. Methodological suggestions are given regarding induction of affect in experimental studies, and assessment of affect. This paper presents the researcher in human performance, automation, and simulation with several resources: a rationale for including affect in the conceptualization of every study; a list of studies that demonstrate affect's role in moderating cognitive and physical performance; references to theoretical models for affect, performance, and workload; references for induction of effect; references for assessment of affect.

Keywords: Affect, mental workload, mood, performance

Affect refers to emotion, including both the current emotional state of an operator, and the emotional traits (that is, the characteristic emotional dispositions) of that operator. Affect involves a subjective filter through which the operator evaluates external conditions. In this paper, we consider the role that affect plays in influencing mental workload and moderating performance. It is our position that (1) affect influences the physical and cognitive aspects of an operator confronting any task, (2) affect thus influences that operator's perception of workload, and (3) affect thus moderates performance on any given task. Thus, all human factors research, including research involving automation and simulation, should involve assessment of affect. To demonstrate the feasibility of this approach, we consider research regarding the influence of affect on cognitive and physical tasks. We then make methodological suggestions.

Affect has been largely ignored in human factors and ergonomics research, and this has been so from the earliest days of the field. Affect did not receive the slightest attention in the early magisterial reviews of the field published by Fitts (1963) and Taylor (1963), despite Taylor's long-ago plea:

> Learning and individual differences sometimes bring about greater changes in operator behavior than do the variables which are preferred by the . . . human engineer. If the engineering psychologist broadened his investigations to include a more deliberate study of social, motivational, individual-difference, and learning factors on psychomotor processes, the goal of proper man-machine system design might be served better. If this were done, conceivably interactions of great importance in optimizing systems – and no less important for psychological theory – would be discovered. (Taylor, 1963, p. 857)

However, it seems that most subsequent research has sided with Fitts, who implied that personality theory was of too little importance to consider at length in his review of engineering psychology (Fitts, 1963, p. 924). There are many possible reasons for this. Much of contemporary human factors and ergonomics psychology is grounded in cognitive psychology, which itself has been criticized for ignoring affect and motivation (Norman, 1980; Pervin & John, 2001). Affect may simply seem too "soft" for a psychological specialization that is rooted in engineering and hard science disciplines.

Nonetheless, abundant research demonstrates that affect influences performance on many tasks. (A theory of the neuropsychological basis of such findings has been proposed by Ashby, Isen, & Turken, 1999.) This has been demonstrated in research involving both cognitive and behavioral performance, often using experimental designs.

Cognitive Performance as Moderated by Affect

Affect influences memory function. Research using an "emotional Stroop task" paradigm demonstrated that, for many people, affect-laden terms delay responses (Williams, Mathews, & MacLeod, 1996). Overall, affective events – particularly negative events – were more recollected by participants, leading to a more rich experience of them in memory (Ochsner, 2000). Memory for incidents demonstrated a "weapon focus"

effect, in which the presence in the incident of a weapon – an affect-laden stimulus that is no stranger to human factors research – inhibited memory for many other details of the incident (Loftus, Loftus, & Messo, 1987; Pickel, French, & Betts, 2003; Steblay, 1992). Memory also exhibits a mood-congruent bias: people with induced happy mood better remembered happy than sad materials, but people with induced sad mood better remember sad than happy materials (Mayer, McCormick, & Strong, 1995).

Affect also influences decision making and problem solving. Depending on other variables, experimentally induced positive affect promoted both risk-prone and risk-averse behavior (Arkes, Herren, & Isen, 1988; Isen & Geva, 1987; Isen, Nygren, & Ashby, 1988; Isen & Patrick, 1983); the same mood induction also increased creativity in problem solving (Estrada, Isen, & Young, 1994; Isen, Daubman, & Nowicki, 1987). In diagnosing liver disease, physicians with induced positive affect integrated diagnostic information quicker than controls; they also showed less distortion or inflexibility in their diagnostic reasoning (Estrada, Isen, & Young, 1997).

Assessments of future probabilities (i.e., risks) are affected by mood. Among healthy people, those with induced happy mood saw themselves as much more likely to experience positive health-related events in the future than those with induced sad moods (Salovey & Birnbaum, 1989). In terms of general life events, participants with induced happy mood reported higher likelihood of positive events occurring, and lower probabilities for negative events occurring, than control participants without induced mood; participants with induced sad mood showed just the opposite pattern, reporting lower likelihood of positive events occurring, and higher probabilities for negative events occurring, than controls (Wright & Bower, 1992; see also Johnson & Tversky, 1983).

Affect influences performance on other cognitively loaded tasks. Induced positive affect increased motivation for and performance on an anagram task (Erez & Isen, 2002). In a word sorting task, participants with induced positive affect used categories more inclusively than controls (Isen & Daubman, 1984). Induced positive mood influenced perceptions of task characteristics and task satisfaction (Kraiger, Billings, & Isen, 1989).

Additionally, affect's role as filter not only applies to external stimuli. Affect and time perception are also interlinked. We all know the saying that 'time flies when you're having fun.' We know this statement to be true, since research has found that positive mood states are associated with subjective underestimates of time passed, while negative affect leads people to report distensions in perceived time passed. (Hornik, 1992).

Finally, affect influences interpersonal functioning. Induced positive affect was associated with use of integrative (rather than contentious) solutions in bilateral negotiations (Carnevale & Isen, 1986).

Physical Performance as Moderated by Affect

Affect seems to moderate performance on tasks that are heavily loaded for strenuous physical activity, though there is a smaller literature on the topic. Among those

attempting to qualify for national wrestling teams in the 1980 Olympics, pre-competition affect was a highly significant predictor of qualifying either for competition or for first alternate status in a given weight class; those wrestlers who so qualified showed lower scores for tension, depression, anger, fatigue, and confusion, while showing higher scores for vigor (Silva, Shultz, Haslam, Martin, & Murray, 1985). Though Lee (1990) found that pre-task mood state did not affect performance on a physical task, the performance was strongly affected by 'psyching up' behaviors, which are strikingly similar to induced affect. This is also in accord with the work of other researchers, who have found that successful elite athletes have positive affect and high levels of self-confidence, and tend to engage in positive self-talk, all prior to competition (Gould, Weiss, & Weinberg, 1981; Highlen & Bennett, 1979; Mahoney & Avner, 1977).

Discussion and Methodological Suggestions

By this point, we believe that we have demonstrated that affect somehow moderates performance (a position we expand on in Koltko-Rivera, Ganey, Dalton, & Hancock, 2004a). Affect may do this by influencing three aspects of a task: (1) its perceived physical characteristics or demands, (2) its perceived cognitive characteristics or demands, and (3) its perceived affective characteristics. (These three types of characteristics define mental workload, that is, the subjective aspects of workload, in a model we are developing; Koltko-Rivera, Ganey, Dalton, & Hancock, 2004b). Thus, human factors scientists should include affect as a moderating variable in their research.

The researcher is always in need of solid theoretical frameworks. Two different types of theories about affect include (1) factorial approaches to personality in general, such as the Five Factor Model (McCrae & Costa, 1999), and (2) circumplex approaches to personality and emotion (Plutchik & Conte, 1997).

Two aspects of method deserve notice: the induction of mood for experimental studies, and the assessment of affect. Many of the studies described above induced mood in research participants; some of the methods used were evaluated by Isen and Gorgoglione (1983). In addition, contemporary simulation scientists have a distinct advantage over earlier researchers. With the advent of large-scale video screens, realistic computer graphics, and sophisticated game engines, the contemporary researcher can induce affect in ways that could only be dreamed of earlier. In terms of assessment, several instruments exist to assess affect, including multi-theoretical instruments (Matthews, Jones, & Chamberlain, 1990; Matthews et al., 1999), Five Factor model instruments (Costa & McCrae, 1992), and instruments appropriate for use in a circumplex model (see multiple chapters in Plutchik & Conte, 1997).

Acknowledgement

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

References

Arkes, H.R., Herren, L.T., Isen, A.M. (1988). The role of potential loss in the influence of affect on risk-taking behavior. *Organizational Behavior & Human Decision Processes, 42*, 181-193.

Ashby, F.G., Isen, A.M., & Turken, A.U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*, 529-550.

Carnevale, P.J., & Isen, A.M. (1986). The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organizational Behavior & Human Decision Processes, 37*, 1-13.

Costa, P.T., Jr., & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Facator Inventory (NEO-FFI): Professional manual.* Lutz, FL: Psychological Assessment Resources.

Erez, A., & Isen, A.M. (2002). The influence of positive affect on the components of expectancy motivation. *Journal of Applied Psychology, 87* 1055-1067.

Estrada, C.A., Isen, A.M., & Young, M.J. (1994). Positive affect improves creative problem solving and influences reported source of practice satisfaction in physicians. *Motivation and Emotion, 18*, 285-299.

Estrada, C.A., Isen, A.M., & Young, M.J. (1997). Positive affect facilitates integration of information and decreases anchoring in reasoning among physicians. *Organizational Behavior & Human Decision Processes, 72*, 117-135.

Fitts, P. M. (1963). Engineering psychology. In S. Koch (Ed.), *Psychology: A study of a science: Vol. 5* (pp. 908-933). New York: McGraw-Hill.

Gould, D., Weiss, M., & Weinberg, R. (1981). Psychological characteristics of successful and nonsuccessful Big Ten wrestlers. *Journal of Sport Psychology*, **3**, 69-81.

Highlen, P.S., & Bennett, B. (1979). Psychological characteristics of successful and nonsuccessful elite wrestlers: An exploratory study. *Journal of Sport Psychology*, **1**, 123-137.

Hornik, J. (1992). Time estimation and orientation mediated by transient mood. *Journal of Socio-Economics, 21*, 209-227.

Isen, A.M., & Daubman, K.A. (1984). The influence of affect on categorization. *Journal of Personality and Social Psychology, 47*, 1206-1217.

Isen, A.M., Daubman, K.A., & Nowicki, G.P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52*, 1122-1131.

Isen, A.M., & Geva, N. (1987). The influence of positive affect on acceptable level of risk: The person with a large canoe has a large worry. *Organizational Behavior & Human Decision Processes, 39*, 145-154.

Isen, A.M., & Gorgoglione, J.M. (1983). Some specific effects of four affect-induction procedures. *Personality & Social Psychology Bulletin, 9*, 136-143.

Isen, A.M., Nygren, T.E., & Ashby, F.G. (1988). Influence of positive affect on the subjective utility of gains and losses: It is just not worth the risk. *Journal of Personality and Social Psychology, 55*, 710-717.

Isen, A.M., & Patrick, R. (1983). The effect of positive feelings on risk taking: When the chips are down. *Organizational Behavior & Human Decision Processes, 31*, 194-202.

Johnson, E.J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology, 45*, 20-31.

Koltko-Rivera, M.E., Ganey, H.C.N., Dalton, J., & Hancock, P.A. (2004a). *Human performance: A comprehensive strategy for research.* Manuscript in preparation, University of Central Florida.

Koltko-Rivera, M.E., Ganey, H.C.N., Dalton, J., & Hancock, P.A. (2004b). *Workload: A comprehensive model.* Manuscript in preparation, University of Central Florida.

Kraiger, K., Billings, R.S., & Isen, A.M. (1989). The influence of positive affective states on task perceptions and satisfaction. *Organizational Behavior & Human Decision Processes, 44*, 12-25.

Lee, C. (1990). Psyching up for a muscular endurance task: Effects of image content on performance and mood state. *Journal of Sport and Exercise Psychology, 12*, 66-73.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus." *Law and Human Behavior, 11*(1), 55-62.

Mahoney, M.J., & Avner, M. (1977). Psychology of the elite athlete: An exploratory study. *Cognitive Therapy and Research, 1*, 135-141.

Matthews, G., Jones, D.M., & Chamberlain, A.G. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology, 81*, 17-42.

Matthews, G., Joyner, L. Gilliland, K., Campbell, S., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf, (Eds.), *Personality psychology in Europe: Vol. 7* (pp. 335-350). Tilburg, The Netherlands: Tilburg University Press.

Mayer, J.D., McCormick, L.J., & Strong, S.E. (1995). Mood-congruent memory and natural mood: New evidence. *Personality & Social Psychology Bulletin, 21*, 736-746.

McCrae, R.R., & Costa, P.T., Jr. (1999). A five-factor theory of personality. In L.A. Pervin & O.P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139-153). New York: Guilford.

Norman, D.A. (1980). Twelve issues for cognitive science. *Cognitive Science, 4*, 1-32.

Ochsner, K.N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing fellings past. *Journal of Experimental Psychology: General, 129*, 242-261.

Pervin, L.A., & John, O.P. (2001). *Personality: Theory and research* (8th ed.). New York: Wiley.

Pickel, K. L., French, T. A., & Betts, J. M. (2003). A cross-modal weapon focus effect: The influence of a weapon's presence on memory for auditory information. *Memory, 11*, 277-292.

Piedmont, R.L. (1998). *The Revised NEO Personality Inventory: Clinical and research applications*. New York: Kluwer/Plenum.

Plutchik, R., & Conte, H.R. (Eds.). (1997). *Circumplex models of personality and emotions*. Washington, DC: American Psychological Association.

Salovey, P., & Birnbaum, D. (1989). Influence of mood on health-relevant cognitions. *Journal of Personality and Social Psychology, 57*, 539-551.

Silva, J.M., Shultz, B.B., Haslam, R.W., Martin, T.P., & Murray, D.M. (1985). Discriminating characteristics of contestants at the United States Olympic Wrestling Trials. *International Journal of Sport Psychology, 16*, 79-102.

Taylor, F. V. (1963). Human engineering and psychology. In S. Koch (Ed.), *Psychology: A study of a science: Vol. 5* (pp. 831-907). New York: McGraw-Hill.

Williams, J.M.G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin, 120*, 3-24.

Wright, W.F., & Bower, G.H. (1992). Mood effects on subjective probability assessment. *Organizational Behavior & Human Decision Processes, 52*, 276-291.

The Peacekeeper: How the Role of the Modern Soldier has Changed and How that Affects Workload

Mark E. Koltko-Rivera
Professional Services Group
Winter Park, Florida

H. C. Neil Ganey
University of Central Florida

Lauren M. Murphy
University of Central Florida

P. A. Hancock
University of Central Florida and
Institute for Simulation and Training

Joseph Dalton
University of Central Florida

Keywords: peacekeeping, military training, workload

Our objective is to make human factors scientists aware of the need for research regarding military peacekeeping. Little research has focused on the soldier as peacekeeper, even though a growing percentage of U.S. soldiers' time is being spent in peacekeeping operations. The combat soldier faces a binary friend/foe discrimination task, and a mission involving the destruction of the enemy and the enemy's warfighting capacity, with relatively less restrictive rules of engagement. The peacekeeper soldier faces a fuzzy friend/foe/non-foe discrimination task, and a mission involving negotiation among disputing factions and preservation of life, with relatively more restrictive rules of engagement. Today's American military forces often must shift between both these roles, resulting in stress that impacts performance in either role. These two roles probably make different workload demands on their operators, suggesting a need for focused research on combat versus peacekeeper workload; some current research programs into these issues are outlined.

> "We were taught how to sneak around these tanks quietly, surprise the enemy, and destroy him in combat. But here we are supposed to stay out of combat by being obvious. To me, it is like teaching a dog to walk backwards." (quote from O'Connor, 1996, p. A-3)

Around the world, various world organizations are banding together in efforts to promote and sustain peace. Over the last 53 years, the United Nations has entered into over 50 peacekeeping operations alone (United Nations Department of Public Information, 2001a). As the United States assumes an increasingly more prominent role in non-combat policing actions around the world, the tasks involved in that policing are falling to our military personnel. Since the end of the Cold War, U.S. presence in what are called Military Operations Other Than War (MOOTW) has increased. These operations are conducted under politically sensitive and complex circumstances. Whether we like it or not, we have become police officer to the world. We are now being called upon to help maintain order. "Peace operations are here to stay and the U.S. Army will be called upon to all corners of the globe to bring stability to chaos, order to anarchy, and peace to conflict" (United States Army Peacekeeping Institute [hereafter USAPI], 1996).

Soldiers are not just required to be warriors anymore. They must now be peacekeepers as well. This comes with a cost. Over 1,200 people have been killed while serving in peacekeeping missions for the United Nations (United Nations Department of Public Information, 2001b). The potential for violent death while peacekeeping means that peacekeeping missions expose soldiers to perhaps at least as much stress as warfighting—if not more so.

Peacekeeping missions are by no means new (Breed, 1998; Moskos, 1976; Segal, 1995). The U.S., however, has not participated in very many peacekeeping missions, due to its involvement in the Cold War (Franke, 2003). We are the new kid on the block of the peacekeeping arena.

What is getting trained?

The problem with this state of affairs is that the training that American soldiers receive is not conducive to peacekeeping actions. American soldiers are excellent warriors, known for their ability to complete their objectives in an efficient and effective manner. However, this does not translate into peacekeeping ability. As Possumato (1999) noted, "Army doctrine and training is focused on how to fight and win wars, not on how to keep peace" (pp. 1-2). A cursory analysis of the numbers of casualties that American troops have sustained since engaging in the first action in Iraq will show that, even though many deaths were incurred during the combat, the number killed while the soldiers have been acting as law enforcement or peacekeeping officers is much greater, almost two times more in hostile situations, almost three times more overall (Directorate for Information Operations and Reports, 2004). It may seem to the casual observer that there is a distinct disconnect here, inasmuch as American soldiers are incurring lethal casualties under supposedly peaceful conditions, when conditions "should be" safer. Part of the reason for this may be that peacekeeping is not seen by the U.S. Army as a separate mission, necessitating the inclusion of specialized preparation into current and future training programs (Possumato, 1999). Rather, peacekeeping is seen as a mission to be

prepared for, "utilizing just enough and just in time procedures, as stated in FM 100-23" (Possumato, 1999, p. iii).

Given the large number of casualties currently being sustained by American troops in Iraq, this would seem to be a perilous procedure. Nonetheless, the American armed forces seems determined to continue in this path. Although consultants have cautioned the Army to not classify peacekeeping as "just extensions of what the Army does" (Possumato, 1999, p.10), and "in spite of the National Security Strategy's assertion that peacekeeping operations will be the most frequent challenge for the U.S. forces for the foreseeable future, BHAAR I (Boznia Herzegovina After Action Review) asserts that the Army still believes an ad hoc approach to training units for such missions is adequate to the task" (Possumato, 1999, p. 4).

An additional risk is that peacekeepers may lose the warfighting skills for which they had originally trained. As Sgt. Felipe Paul put it, "I'm a tanker [tank operator], that's what I do, been one for 14 years. But let me tell you, those skills are perishable. You got to use them, and all I'm doing here is checking people's driver's licenses" (quoted in O'Connor, 1996, p. A-3). As a tank commander said of his men, "my guys are great at driving three Bradleys (tanks) down a road in a straight line and setting up an unconcealed observation post, but when it comes to attacking a position, or holding a piece of terrain against an assault, that's where we'll need work" (Boccardi, quoted in O'Connor, 1996, p. A-3). These comments rings true with many Army officers, who acknowledge that peacekeeping troops will require re-training in their warfighting skills prior to their being sent to their next mission: "In a crisis we would still need about two months of retraining to get everyone coming back now ready" (Col. Goff, quoted in O'Connor, 1996, p. A-3).

The task of peacekeeping

Mission types engaged in by military units fall into several general categories. These categories are combat, peace support (which includes peacekeeping), and humanitarian efforts. With the increase in complexity that we have witnessed in contemporary operations, some missions contain aspects of more than one of those categories (Britt & Adler, 2003). Even within the category of peace support, there exist different functions. Peacekeeping missions, as the public understands them, support the building of peace in a region through non-violent aid. Peace enforcement, however, gives authorization for the use of force to encourage parties to sustain peace.

An example of a mission that had dynamic shifts in mission type was the Army operation in Mogadishu. The Army's presence in Somalia was to prevent the local warlords from intercepting aid that was being provided for the population of Mogadishu. This mission was drastically changed when the local militia attacked U.S. forces, using many different methods that the American soldiers were unaccustomed to, including surrounding themselves with women and children to serve as shields for them. The American forces were caught off guard by these tactics and were unprepared to meet them. "For the forces on the ground (in the Somalian conflict), transitioning between operational assignments meant continuous change in the mission and its objectives"

(Franke, 2003, p. 33). As peacekeepers, soldiers must understand that, although they may be required to avoid the use of force as much as possible, they should be prepared to use all necessary force, should the tables turn for the worse. However, they also must understand how to use good judgment in the use of that force, given that they are there primarily to keep the peace. "Every [peacekeeping] soldier must be aware that the goal is to produce conditions that are conducive to peace and not to the destruction of an enemy" (Possumato, 1999, p. 5). Thus, peacekeepers are required to conduct a balancing act.

"The challenge of these missions lies not only in the fluid nature of the operational environment but also in the contrast between the goals of peacekeeping and the expectations and training of combat-ready military forces" (Britt & Adler, 2003, p.6). Britt and Adler (2003) have presented a framework of hypothesized factors affecting the performance of the peacekeeper, breaking the stressors experienced by the peacekeeper into three basic categories of stress: factors of the specific mission, factors relevant to the peacekeeper's unit, and factors relevant to each peacekeeper. Factors relevant to the mission include the rules of engagement, presence (or absence) of a peace treaty, and presence (or absence) of public support of the peace. Unit cohesion and morale and pre-deployment training would fall under factors relevant to the unit. These factors match up very closely to the concepts of environmental, internal, and task-based workload (Hockey & Hamilton, 1983), as they include aspects of the mission (task), their unit (environment), and themselves (internal).

The case for specialized peacekeepers

Some countries have recognized that what makes a good combat soldier is not necessarily what makes a good peacekeeper. These governments have therefore created separate forces for these roles (Wisher, 2003). Others, like the United States, select forces and give them pre-deployment training which, as stated earlier, may or may not conflict with the training that they have previously received from the military.

The missions engaged in by peacekeepers primarily differ from combat missions in the fact that there is no obvious enemy to focus on (Franke, 2003). Additionally, in many cases, peacekeepers lack the clear mission objectives and 'field of battle' that is customary to war. Warriors, operating under less restrictive rules of engagement, are able to prevail against clearly defined hostile forces, through the application of violence. Peacekeepers, however, must utilize diplomacy and avoid violence in the course of seeking peaceful resolutions to situations. Thus, we are presented with a case where soldiers are being tasked with jobs wherein they must assume different, even conflicting, roles (Diehl, 1988; Kutter, 1986).

In performing research on the peacekeeper, we would attempt to aid soldiers selected for MOOTW to maximize performance under dynamic conditions. We would seek ways to mitigate the stress, workload, and fatigue that is experienced by peacekeepers as they carry out their operation. It has been said that soldiers (as well as others who perform in extreme environments) experience hours of boredom interrupted by moments of terror. This is the case, not only for warriors, but for peacekeepers as well. A dynamic change in the state of a scenario (i.e., from peacekeeping to warfighting) may

cause many operators to falter, as they struggle to adjust their schema of the situation to match with their current perceptions. As Castro (2003) suggests, we need to establish what the ideal states for optimal performance in peacekeeping missions are, in order to use them as criteria for treatment outcomes.

The tasks of the peacekeeper and the warrior may differ in terms of their somatic, cognitive, and affective features and are worthy of extensive research. The issue here, at an abstract level, involves differences in workload, which is the focus of the Workload Theory and Measurement research program within the Multiple University Research Initiative for Operator Performance Under Stress (MURI-OPUS) laboratory at the University of Central Florida (UCF; Koltko-Rivera et al., 2004). In addition to describing the difference between peacekeeper and warrior workload, we will seek methods to mitigate the stress and workload associated with the dynamic characteristics of the task that peacekeepers experience.

Other activities within the UCF MURI-OPUS laboratory are relevant to this task, as well. Galantino (2003) has described how the peacekeeper is made to change from a binary friend/foe mentality to a fuzzy friend/foe/non-foe classification model. This is directly related to the work that Peter Hancock and his colleagues at the UCF MURI-OPUS laboratory are doing, in developing fuzzy signal detection theory (Hancock, Masalonis, & Parasuraman, 2000; Parasuraman, Masalonis, & Hancock, 2000). By analyzing the choices these warriors *cum* peacekeepers are making, as well as measures of the workload that the soldiers are experiencing, we can determine the extent to which soldiers on peacekeeping missions are subjected to different workloads than their warfighting counterparts. This may provide compelling evidence for the peacekeeper position to be reassessed as a trained position within the U.S. armed forces.

In this paper, we have tried to bring to light a problem that American soldiers are facing in their roles as peacekeepers. We invite all human factors scientists to join with us in investigating the nature of the challenges they face, and in making recommendations to mitigate these stresses. Such research can help peacekeeping forces of all nations, and help to build a safer world.

Acknowledgement

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

References

Breed, H. (1998). Treating the new world disorder. In H. J. Langholtz (Ed.), *The psychology of peacekeeping*. (pp. 111-128). Westport, CT: Praeger.

Britt, T.W., & Adler, A.B. (2003). The psychology of the peacekeeper: An introductory framework. In T.W. Britt and A.B. Adler (Eds.), *The psychology of the peacekeeper* (pp. 3-10). Westport, CT: Praeger.

Castro, C.A. (2003). Considerations when conducting psychological research during peacekeeping missions: The scientist and the commander. In T.W. Britt and A.B. Adler (Eds.), *The psychology of the peacekeeper* (pp. 11-28). Westport, CT: Praeger.

Diehl, P.F. (1993). *International peacekeeping*. Baltimore and London: Johns Hopkins University Press.

Directorate for Information Operations and Reports (2004) *Operation Iraqi Freedom – Military Deaths*. Retrieved February 1, 2004, from http://web1.whs.osd.mil/mmid/casualty/castop.htm

Franke, V.C. (2003). The social identity of peacekeeping. In T.W. Britt and A.B. Adler (Eds.), *The psychology of the peacekeeper* (pp. 31-52). Westport, CT: Praeger.

Galantino, M.G. (2003). Work motivation and the peacekeeper. In T.W. Britt and A.B. Adler (Eds.), *The psychology of the peacekeeper* (pp. 111-126). Westport, CT: Praeger.

Hancock, P.A., Masalonis, A.J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science, 1*, 207-230.

Hockey, G.R.J., & Hamilton, P. (1983) The cognitive patterning of stress states. In G.R.J. Hockey (Ed.) Stress and fatigue in human performance, pp. 331-362. New York: John Wiley and Sons.

Koltko-Rivera, M.E., Ganey, H.C.N., Dalton, J., & Hancock, P.A. (2004). *Workload: A comprehensive model*. Manuscript in preparation, University of Central Florida.

Kutter, W.D. (1986). *Operational guidelines for U.S. peacekeeping commanders.* Carlyle Barracks, PA: U.S. Army War College.

Moskos, C. (1976). *Peace soldiers: The sociology of a United Nations military force.* Chicago: University of Chicago Press.

O'Connor, M. (1996, December 13). Does keeping the peace spoil G.I.'s for war? *New York Times,* p. A-3.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, *42*, 636-659.

Possumato, D.P. (1999) *Should the U.S. Army establish a peacekeeping training center?* Carlyle Barracks, PA: U.S. Army War College.

Segal, D.R. (1995). Five phases of United Nations peacekeeping: An evolutionary typology. *Journal of Political and Military Sociology*, *23*, 65-79.

United Nations Department of Public Information. (2001a). *United Nations peace operations in 2001.* Retrieved January 22, 2004, from http://www.un.org/Depts/dpko/dpko/pub/year_review01/index.html

United Nations Department of Public Information. (2001b). *United Nations peacekeeping: In the service of peace.* Retrieved January 22, 2004, from http://www.un.org/Depts/dpko/dpko/home_bottom.htm

United States Army Peacekeeping Institute (1996). *Bosnia Herzgovina After Action Review I.* Retrieved February 3, 2004, from http://www.fas.org/man/dod-101/ops/docs/bosrep2a.htm

Wisher, R.A. (2003. Task identification and skill deterioration in peacekeeping operations. In T.W. Britt and A.B. Adler (Eds.), *The psychology of the peacekeeper* (pp. 91-110). Westport, CT: Praeger.



University of Central Florida



Menu Design Considerations for Rapid Decision-Making Simulations

William Y. Pike, Jessica Helmick-Rich, Kelly Burke,
Adams Greenwood-Ericksen, Kristy A. Bradley

Agenda

- Introduction
- Problem
- Hypothesis
- Method
- Issues
- Expected Contributions

Introduction

- Requirement for menus vs keyboard control
  - Limited usage training systems
  - Rapid decision making
  - Multiple commands to recall
- Need for optimally designed menus

Introduction, cont.

- Rapid Decision Trainer (RDT)
  - Target audience: Infantry Officer Basic Course
  - Replicate live-fire exercises
  - Force students to make rapid decisions
  - Limited number of scenarios
  - Ease of use outweighs other design criteria
    - Menu-driven interface

Introduction, cont.

- Previous Research on Menu Design
  - Color coding
    - Tullis, 1981; Yeh & Wickens 2001
  - Grouping
    - Kahneman & Henick, 1977; Winzenz, 1972
  - Display density
    - Brown & Monk, 1975; Triesman, 1982

Introduction, cont.

- Divided Attention (Wickens, 1992)
  - Lack of previous research regarding effects on menu design
  - Importance of investigating divided attention and menu design

Problem

- Previous pilot study indicated significant differences from expected (based on literature) results
- Findings might have strong implications not only for menu design but also for the way in which perceptual information is processed

Hypothesis

- For the color coded menu, it is hypothesized that:
  - Participants will perform better on the primary task in a color coded menu design as compared to a non-color coded control design
  - When the secondary task is presented, a greater decrement in performance on the primary task will be observed in a non-colored design as compared to a color-coded design

Hypothesis, cont.

- For the grouped display menu, it is hypothesized that:
  - Performance will be better in a menu display grouped by command type as compared to a non-grouped control display, but performance will be degraded in the presence of a secondary task

Hypothesis, cont.

- For the density display menu, it is hypothesized that:
  - Performance of participants will be better on the menu design with lower local and overall densities than the performance of participants using the control menu design
  - Performance of participants using the low density menu design will be degraded in the presence of a secondary task

Method

- Modeled current RDT menus with Inquisit (experimental design software)
- Developed 4 menus:
  - Control (existing menu)
  - Color-coded
  - Grouped-display
  - Density-display
- Modeled typical RDT tasks
  - e.g., shift fire signal during assault

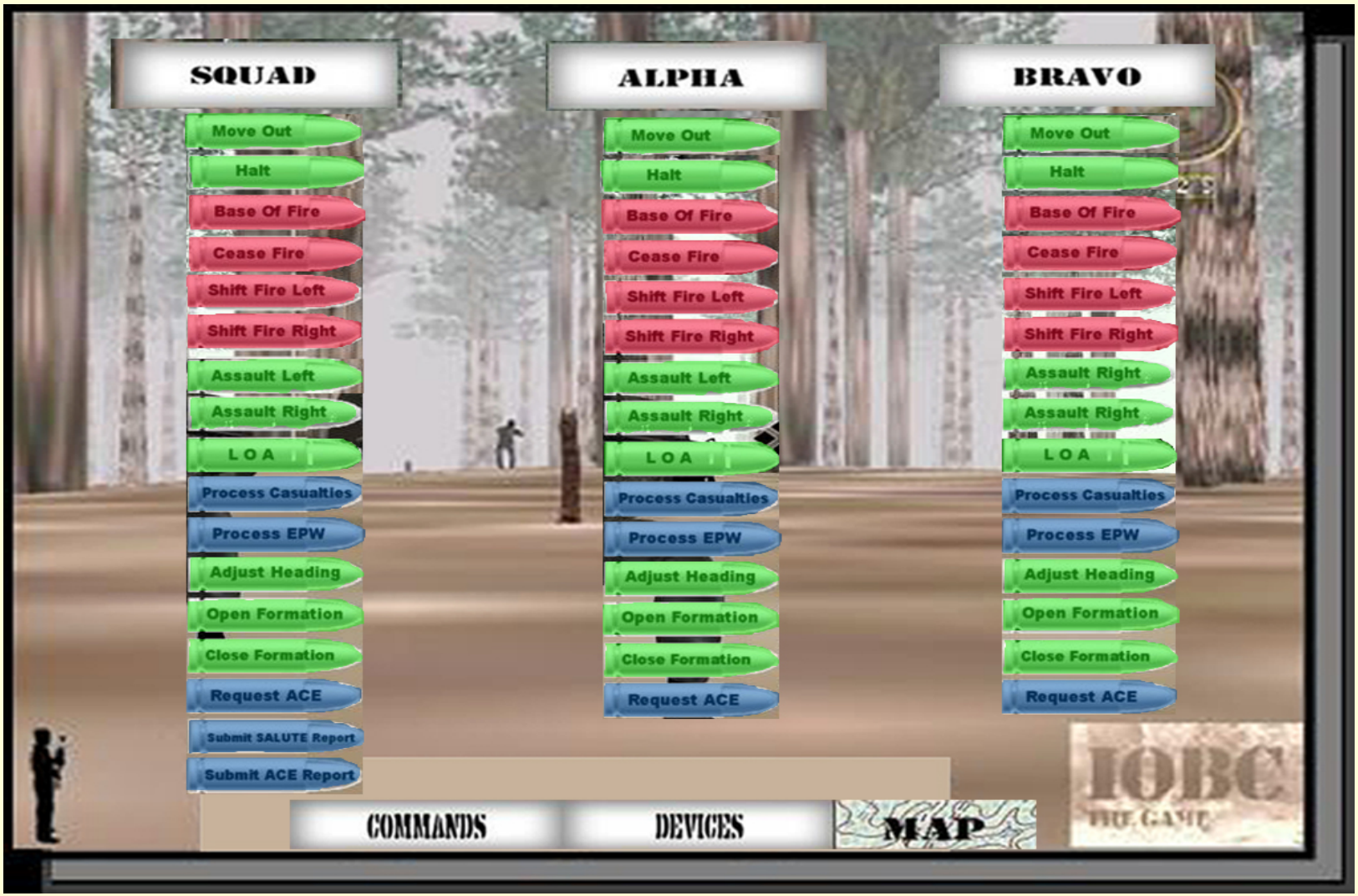
Method, cont.

- Inquisit
  - Demographics
  - Practice with primary and secondary tasks
  - Practice with NASA-Task Load Index (TLX)
  - 8 blocks, 10 trials each
  - Randomized presentation of 4 menu configurations and presence/absence of secondary task
  - TLX between each block

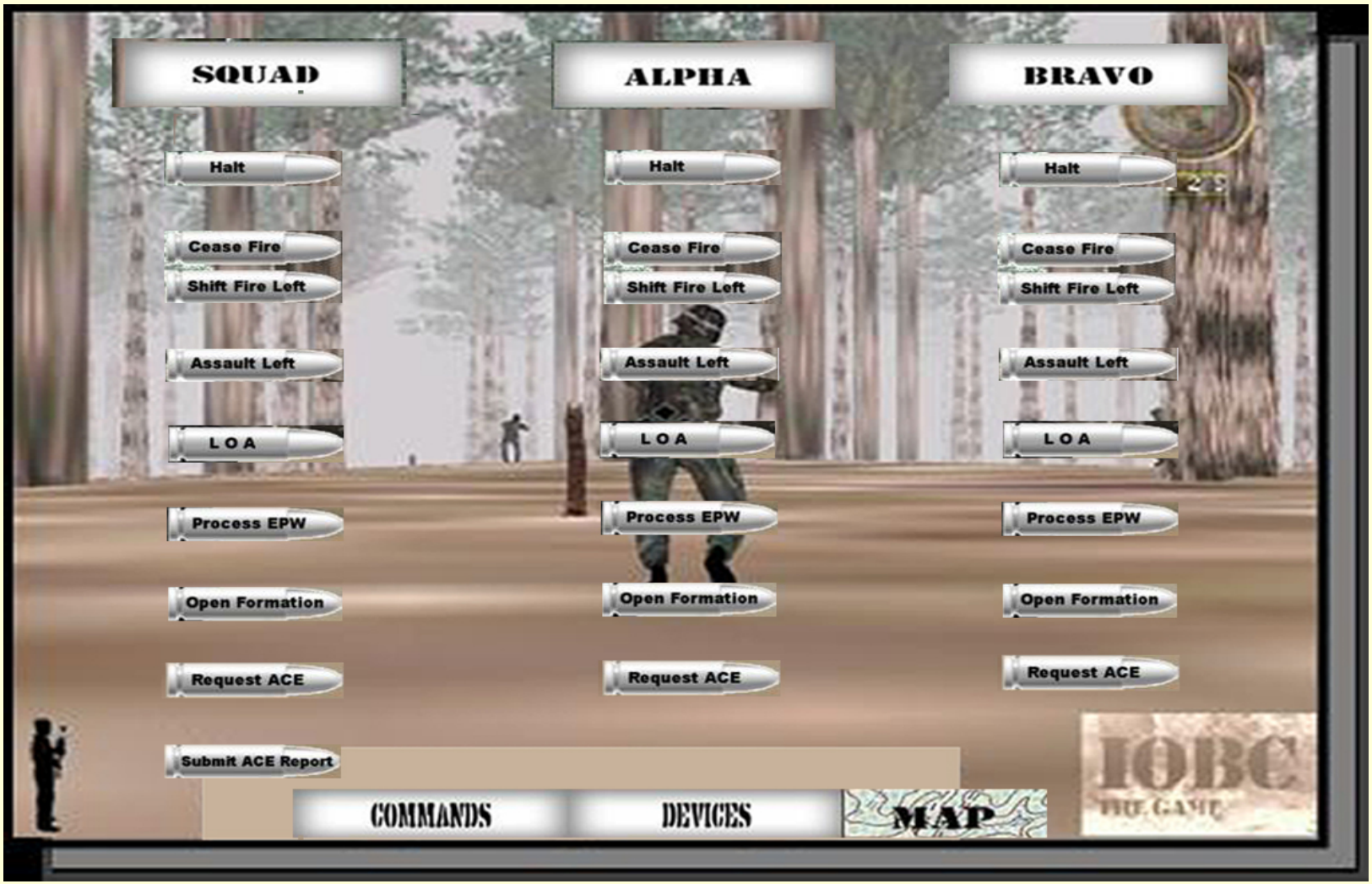
Control Menu



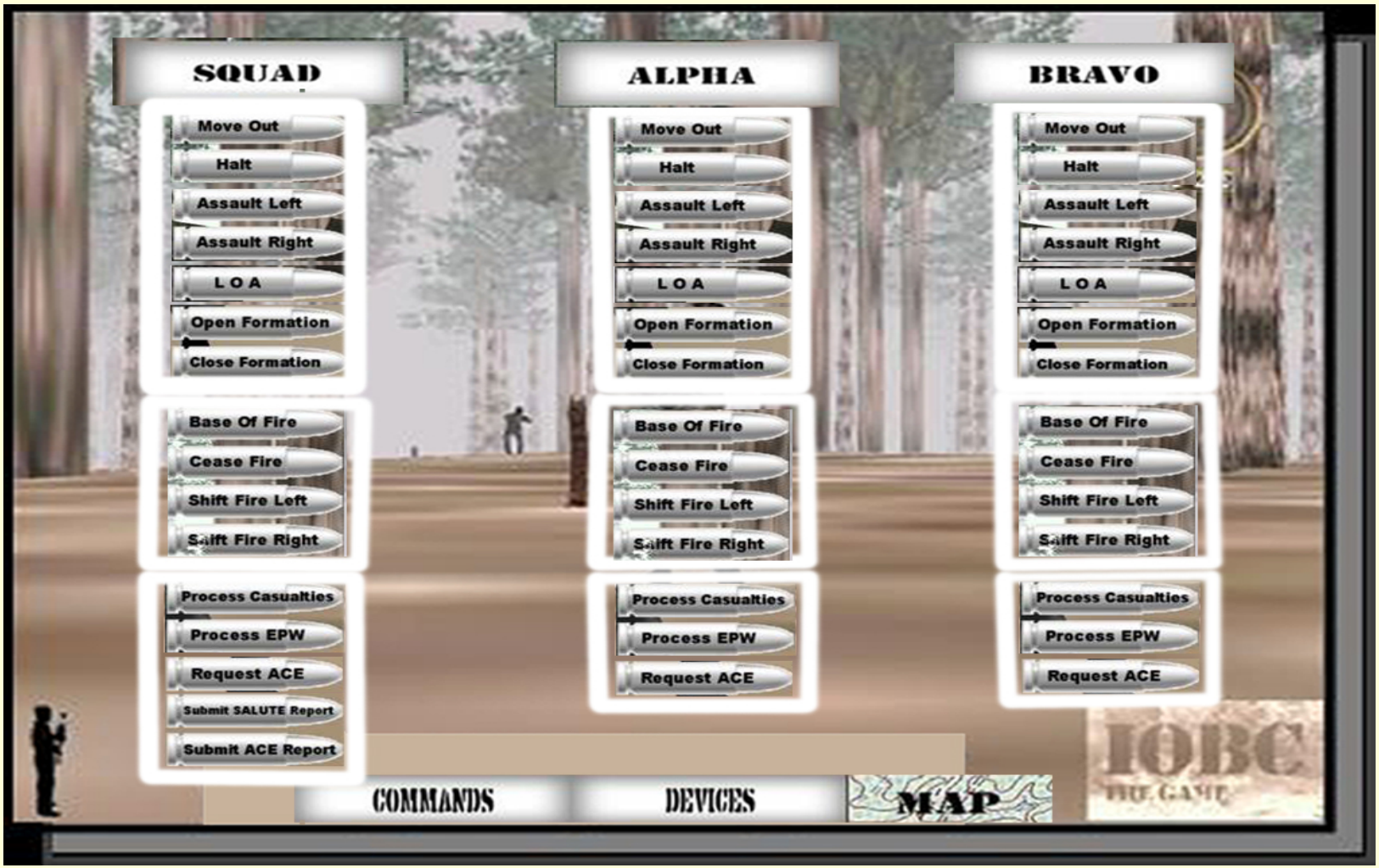
Color Grouping Menu



Density Menu



Grouping Menu



Issues

- Resolved
  - Possible methodological artifacts
    - Primary task delay between presentations of stimuli
    - Secondary task stimuli too difficult to discriminate
  - Insufficient sample size to draw effective inferences

Expected Contributions

- Result in better guidelines for designers
  - Specifically, menu-heavy training applications
- Advance research into menu color-coding, grouping, & density with respect to added stressors
- Contribute to our understanding of information processing under stress

Efficacy of Transfer in Simulation-Based Training: Implications for Stress Exposure Training

Jennifer M. Ross, James L. Szalma, & Peter A. Hancock

Psychology Department
And Institute of Simulation and Training
University of Central Florida
Orlando, FL

Abstract

Transfer of training from simulators to the real-world has recently come under investigation as the generalizability of task-specific training has come into question. New hypotheses recommend that, to ensure effective performance under stress in real-world environments, one should supplement skill-based training with stress exposure training. Stress exposure training has further benefits in that it may serve as a more generalizable form of training and transfer across tasks and stressors. The impact of improving performance and reducing perceived stress and workload is of vital importance to many military operations, especially in high technology and high workload situations such as the Landwarrier or Unmanned Arial Vehicles (UAVs), in which mistakes are costly in terms of economics as well as life. In this article the premise of and future implications of Stress Exposure Training are discussed.

Simulation Training

Simulation has been successfully employed in military and commercial sectors to train individuals (e.g., pilots, soldiers, doctors) with the objective of increasing their performance accuracy in real-world situations. The primary method of performance improvement in these contexts is to provide simulator-based practice of situations and conditions similar to those expected to be encountered in a real-world Stress Exposure Training. However, the extent to which simulator-based training transfers to real-world application has recently been questioned (Driskell, Johnston, & Salas, 2001). New theories of simulator training challenge traditional task-based simulation training, claiming that skill acquisition is highly task-specific and not readily transferable to different tasks and situations (Sims & Mayer, 2002). This limitation decreases the applicability of simulator training transferring to performance in a real-world Stress Exposure Training, which is highly complex, dynamic, and virtually impossible to anticipate all potential task/accident scenarios that may occur. In a review by Ivancevich and colleagues (1990), recommendations were made that training programs should be designed to address performance outcomes associated with specific stressors. Indeed, such disasters as Three Mile Island, Chernobyl, and the USS Vincennes have underscored the importance of developing training interventions to offset the impact of stressors on complex cognitive tasks (Johnston & Cannon-Bowers, 1996).

Stress Exposure Training

One avenue of increasing the generalizability of simulator-based training has been through empirical work which has been conducted on Stress Exposure Training (Klepac, Hauge, Dowling, & McDonald, 1981; Saunders, 1993; Saunders, Driskell, Johnston, Salas, 1996). Considerable evidence has shown that stressors cause decrements in attitudes and performance, which are related to outcomes that are emotional, behavioral, or physiological in nature (Greenhaus & Parasuraman, 1987). Stress Exposure Training works by training individuals on factors that are more generalizable than task skills, such as controlling perceived stress and workload by providing intervention information on stressors, sensory information on physiological and emotional reactions to stress, and procedural information on how stress effects task performance. These three phases of stress exposure training will be further discussed later in the paper.

History Stress Exposure Training

Since the inception of stress exposure from Wolpe's work (as cited in Johnston & Cannon-Bowers, 1996) and eventually cognitive/behavioral stress-coping training in the early 1970s, the application of stress exposure training has generally been clinical, and applied chiefly to uses, as alleviating physical pain, anxiety, depression, and anger (Deffenbacher & Suinn, 1988). Gradually, stress-coping training was expanded to include training to a number of occupations (e.g., nurses, police officers, oil-rig trainees), and to enhance athletic performance (Johnston & Cannon-Bowers, 1996). In a meta-analysis by Saunders, Driskell, Johnston, and Salas (1996; See also Saunders, 1993) it was demonstrated that this type of training is an effective means for reducing state anxiety, reducing skill-specific anxiety, and enhancing performance under stress.

Johnston & Cannon-Bowers (1996) argues that there is not sufficient research to suggest that training developers adopt the strategies used in cognitive/behavioral stress coping training programs. This was due to two reasons: first traditional stress exposure research had too narrow a focus (clinical and sports), and second, little work had been done to integrate research findings into a conceptual framework to guide the development of incorporating stressors into training and examining coping skills training for complex cognitive tasks. Therefore, they advocated the use of the term Stress Exposure Training to extend the focus of 'cognitive/behavioral stress coping' beyond its original clinical domain.

Stress Exposure Training has three main objectives, 1) to build skills that promote effective performance under stress, 2) to build performance confidence, and 3) to enhance familiarity with the stress environment. According to Meichenbaum (1985) these three objectives can be accomplished by employing two components: 1) stress-coping training – features developing skills that reduce potential cognitive and psychomotor performance deficiencies resulting from specific stressors (Driskell & Salas, 1991) and 2) instructional design – features a process of gradual exposure to realistic stressors that enhance learning of coping skills, and is based on basic principles of training design for skills acquisition. The two integrated components have resulted in a three-phase process (See Table 1; Meichenbaum, 1985).

The Three Phases of Stress Exposure Training

Phase 1. The first phase in stress exposure training typically involves a discussion of common reactions people have to specific stressors they encounter, with the objective to help the individual understand that they will be able to stop the negative thoughts and behaviours that contribute to stress. It is expected that as the individual's anxiety is reduced, perceived control increases, and consequently ones efficacy in dealing with problem stressors is improved (Johnston & Cannon-Bowers, 1996). Interestingly, violations of expectations have been found to be universally detrimental to performance, even when participants' expectations are violated by their situation being better than expected (Keinan & Friedland, 1996). The effect of credible prepatory information on performance has been found to result in reduced subjective stress, enhanced task confidence, and fewer errors (Inzana, Driskell, Salas, & Johnston, 1996).

Phase 2. The second phase focuses on learning stress coping skills through practice and feedback. The objective is to train the individual to maintain an awareness of stress reactions in order to invoke appropriate skills to reduce stress. The act of maintaining an awareness of thoughts and actions is referred to as "metacognition" (Glaser & Bassok, 1989). Therefore, metacognitive skills; such as, thought restructuring (replacing negative thoughts and reactions that are triggered by a stressor with positive coping thoughts and reactions), problem solving (reduce task performance errors), & physiological control (involves deep breathing and muscle relaxation methods to calm physiological reactions to stressful encounters) support the execution of competent performance.

Phase 3. The third phase involves practicing the coping skills in a setting that simulates or reproduces the problem stressors. Ways of coping with stress are often classified into three areas: task-focused coping, emotion-focused coping, and avoidant-focused coping (Lazarus & Folkman, 1984). Task-focused coping circumvents negative affect through cognitive and behavioural activity that minimizes or modifies the stressor. In this case the individual focuses on dealing with the task and develops strategies aimed at accomplishing the goals of the task (Thropp, Szalma, Ross, & Hancock, 2003). Emotion-focused coping consists of attempts to alleviate or eliminate emotions elicited by a stressor directly, with little attention paid to the characteristics of the situation or to the nature of the threat it poses (Keinan & Friedland, 1996). While, emotion-focused coping has been found to be effective in preparing individuals to cope with stressors that they cannot control and which are brief in duration, it has proved less effective/relevant for training people to withstand and cope with combat stressors. This more prolonged and controllable situations requires mixed or problem-focused coping. The third coping strategy, avoidant coping, involves diverting one's attention from the stressful situation (see also Cox & Ferguson, 1991). Expected outcomes from this phase are reduced anxiety, increased perceived efficacy toward performance, and improved cognitive and psychomotor performance. Further research involves examining phased-training procedures that divide task acquisition without stress exposure from practice of newly acquired skills under stress. Other methods of skill practice under stress include graduated-intensity training (Keinan & Friedland, 1996).

Table 1: Stress Exposure Training Design (as cited from Johnston & Cannon-Bowers, 1996).

Stress Exposure Training Design			
	Phase 1	Phase 2	Phase 3
	Presentation of requisite knowledge	Skill practice with feedback	Skill practice with stressors
Objectives	• Knowledge of typical reactions to stressors.	Develop metacognitive skills • Positive coping thoughts and behaviors • Use relaxation techniques to calm physiological reactions • Develop cognitive skills • Use problem solving skills	• Use phase 2 skills while exposed to stressors.
Outcomes	• Increased perceived efficacy in dealing with stressors.	• Reduced Negative attitudes toward self and stressors. • Increased use of positive thoughts and behaviors. • Reduced blood pressure, heart rate, and increased psychomotor steadiness. • Successful coping skill performance.	• Reduced anxiety. • Increased efficacy. • Successful application of skills while exposed to stressors. • Improved cognitive and psychomotor performance under stress.

Theory behind Stress Exposure Training

One can examine the mechanism of Stress Exposure Training through the transactional model of stress. In this view stress occurs when the perceived demands of the situation tax or exceed the perceived resources of the individual to meet those demands (Lazarus & Folkman, 1984). Based on this theory, Meichenbaum (1993; as cited in Johnston & Cannon-Bowers, 1996) has proposed that Stress Exposure Training provides the skills that should reduce the imbalance between the demands of the stressful situations and the individual's coping resources. The benefits of decreasing this discrepancy have a far-reaching impact on stress and performance.

One can further, examine the relationship between stress and performance by referring to the Hancock and Warm dynamic model of stress and sustained attention (1989). In this model, performance is directly influenced by workload and stress factors, where stress is integrally related to environment (Hockey, 1983), appraisal/coping mechanisms of the exposed individual (Lazarus, 1966), and the general response of the physiological system (Selye, 1956; see Figure 1). As these factors combine and lead to debilitative levels of hyper- or hypo-stress the operative state begins to degrade first psychologically, in increased perceived stress and workload, and then physically through impaired performance. Support for this view has been found across two meta-analyses. In a meta-analysis, by Driskell and colleagues, the effects of stressors on self-reported stress and performance was examined, and it was found that people were more likely to report "feeling stressed" before showing evidence of performance impairment (as cited in Saunders, 1993). In a separate meta-analysis conducted by Saunders (1993) moderate magnitude effects were found for stress inoculation training (a similar three phase stress exposure training methodology) for state and skill based anxiety, but only small magnitude effects for performance. From this Saunders drew the conclusion that subjective reports of stress were more sensitive to the effects of stress inoculation training than performance; that is, reports of lower perceived stress would occur before signs of performance improvement.

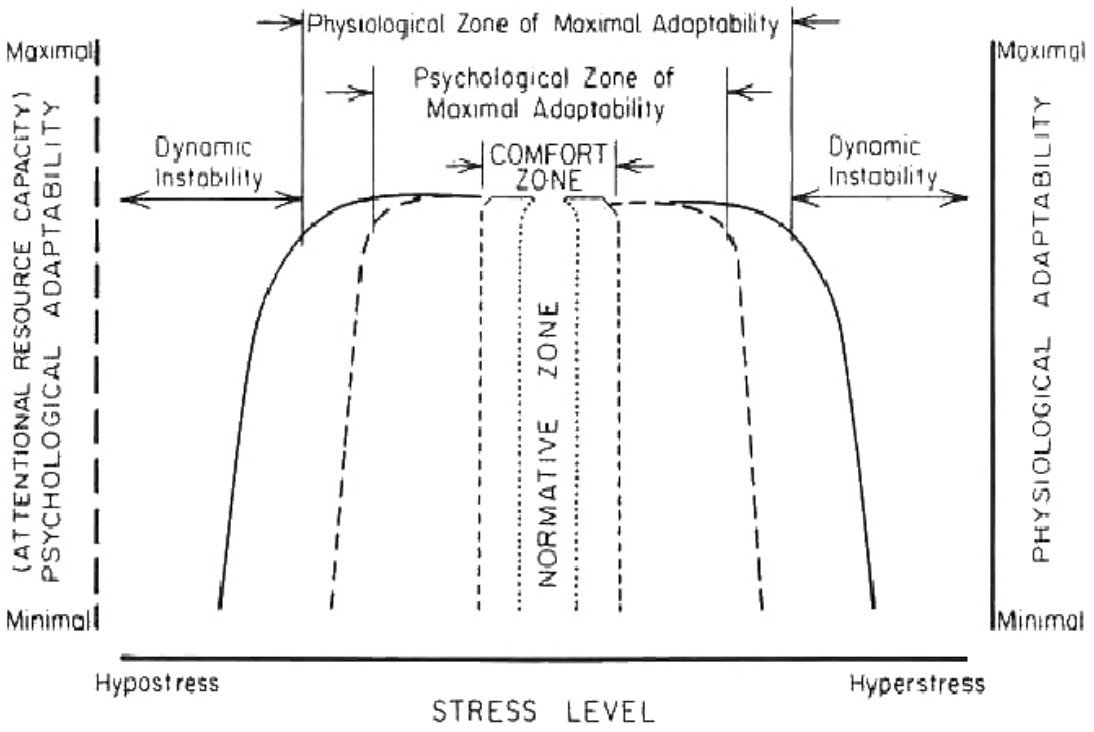


Figure 1: Hancock and Warm dynamic model of stress and sustained attention (as cited from Hancock & Warm, 1989).

New Directions

Research examined in the meta-analysis conducted by Saunders and colleagues (1993, 1996) has supported the view that stress impacts observers psychological before performance begins to degrade. While, these hypotheses support the Hancock and Warm model of dynamic stress and sustained attention, it would be critical to examine the degree to which task-based environmental factors and coping strategies as moderators of Stress Exposure Training efficacy. It is further imperative to investigate limits of transfer of training between field and laboratory training. Studies are currently under investigation to examine the effects of spatial-temporal perceptions on performance and perceived stress/workload with and without stress exposure training. The effects of individual differences in coping strategy will be used in regression to examine the moderating factors of this variable on Stress Exposure Training efficacy. Finally, while Saunders (1993) found Stress Exposure Training to beneficially affect performance in the lab and the field, they did not examine the transfer of Stress Exposure Training from a lab to a field task. Indeed, as simulations used in training environments become more realistic the question of specificity becomes more important. That is as realism is improved; this could lead to lower generalizability and hampered transfer of training. These factors represent new avenues of research that must be examined to fully develop the construct of Stress Exposure Training. Figure 2 represents a proposed model to test Stress Exposure Training as related by the moderating variables of environmental factors (i.e., spatial or temporal task demands), individual differences in coping style, and training transfer.

Conclusion

The generalizations of training effects from one task/stressor to another are of critical concern. Future research must examine how Stress Exposure Training principles transfer from the training to the operational environment, and which factors facilitate or inhibit this transfer. Clearly, the cost of training might be prohibitive if research shows that the effectiveness of stress training is limited to specific stressor-task combinations.

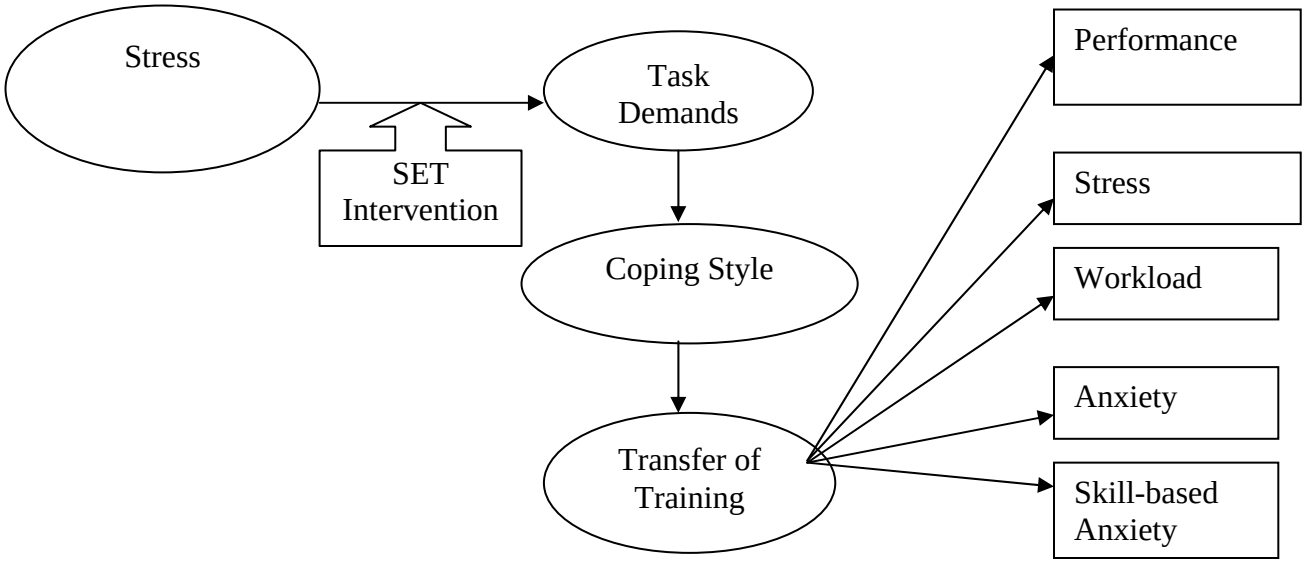


Figure 2: Hypothesized relationship between Stress Exposure Training and Moderators of environmental task demands, individual differences in coping style, and training transfer.

References

Cox, T & Ferguson, E (1991) Individual difference, stress and coping. In Cooper , C and Payne, L (Eds) *Personality and stress: individual differences in the stress process.* (pp 7-29). Chichester, UK: Wiley & Sons.

Deffenbacher, J.L., & Suinn, R.M. (1988). Systematic desensitization and the reduction of anxiety. *The Counseling Psychologist, 16(1),* 9-30.

Driskell, J.E., & Johnston, J.H. (1998). Stress Exposure Training. In J.A. Connon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 191-217). Mahwah, N.J. : Lawrence Erlbaum Associates.

Driskell, J.E., Johnston, J.H., & Salas, E. (2001). Does stress training generalize to novel Stress Exposure Trainingtings? *Human Factors, 43(1),* 99-110. Mahwah, N.J. : Lawrence Erlbaum Associates.

Driskell, J.E., & Salas, E. (1991). Overcoming the Effects of Stress on Military Performance: Human Factors, Training, and Selection Strategies. In G. Reuven & A.D. Mangelsdorff (Eds.), *Handbook of military psychology* (pp. 183-193). NewYork, NY: John Wiley & Sons.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review Psychology, 40,* 631-666.

Greenhaus, J.H., & Parasuraman, R (1987). A work/nonwork interactive perspective of stress and its consequences. Special issue: Job stress: From theory to suggestion. *Journal of Organizational Behavior Management, 8,* 37-59.

Hancock, P.A., and Warm, J.S. (1984). A dynamic model of stress and sustained attention. *Human Factors, 31(5),* 519-537.

Hockey, R. (1983). *Stress and fatigue in human performance.* Chichester, UK: Wiley.

Inzana, C.M., Driskell, J.E., Salas, E., & Johnston, J.H. (1996). Effects of preparatory information on enhancing performance under stress. *Journal of Applied Psychology, 81(4),* 429-435.

Ivancevich, J.M., & Matteson, M.T., Freedman, S.M., & Philips, J.S. (1990). Worksite stress management interventions. *American Psychologist, 45,* 252-261.

Johnston, J.H., & Cannon-Bowers, J.A. (1996). Training for Stress Exposure. In E. Salas (Eds.), *Stress and human performance* (pp. 223-256). Hillsdale, NJ: Erlbaum.

Keinan, G., & Friedland, N. (1996). Training Effective Performance Under Stress: Queries, Dilemmas, and Possible Solutions. In E. Salas (Eds.), *Stress and human performance* (pp. 257-277). Hillsdale, NJ: Erlbaum.

Klepac, R.K., Hauge, G., Dowling, J., & McDonald, M. (1981). Direct and generalized effects of three components of stress inoculation for increased pain tolerance. *Behavior Therapy, 12(3),* 417-424.

Lazarus, R.S. (1966). *Psychological stress and the coping process.* New York: McGraw-Hill.

Lazarus, R.S., & Folkman, S. (1984). *Stress appraisal and coping*. New York: Springer.

Meichenbaum, D. (1985). *Stress Inoculation Training.* New York, NY: Pergamon Press.

Saunders, T.L. (1993). *The effect of stress inoculation training on anxiety and performance.* Unpublished master's thesis, University of Central Florida, Orlando, Florida, United States.

Saunders, T., Driskell, J.E., Johnston, J.H., & Salas, E. (1996). The effect of stress inoculation training on anxiety and performance. *Journal of Occupational Health Psychology, 1(2),* 170-186.

Selye, H. (1956). *The stress of life.* New York: McGraw-Hill.

Sims, V.K., & Mayer, R.E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied cognitive psychology, 16(1),* 97-115.

Thropp, J.E., Szalma, J.L., Ross, J.M., & Hancock, P.A. (2003). Individual differences in dispositional pessimism, stress, and coping as a function of task type. *Proceedings of the Human Factors and Ergonomics Society, USA, 47,* 1073-1077.

Fuzzy Signal Detection Theory: Basic and Applied Research

J.L. Szalma, S.C. Stafford., T. Oron-Gilad., L.L. Murphy, & P.A. Hancock

Abstract

A recent advance on Signal Detection Theory (SDT) (see Swets, 1996) promises to enhance the measurement of performance in complex real world domains. This development, Fuzzy Signal Detection Theory (FSDT; Hancock, Parasuraman, & Masalonis, 2000; Parasuraman, Masalonis, & Hancock, 2000), combines traditional SDT with Fuzzy Set Theory to extend analysis beyond the traditional crisp, categorical decision making model. FSDT allows for events to simultaneously be in more than one state category (e.g. signal and non-signal), so that the stimulus and response dimensions can be continuous rather than categorical. Consequently, FSDT can be employed in applied settings where the degree to which an event is a signal for detection may vary. This paper presents, albeit briefly, the tenets of FSDT and ways it can be applied to the problems faced by those designing real-world detection systems.

Elements of Fuzzy Signal Detection Theory

Events can belong to the set "signal" (**s**) to a degree ranging from 0 to 1

Events can belong to the set "response" (**r**) to a degree ranging from 0 to 1

After defining these sets FSDT measures can be derived. (Parasuraman, Masalonis, & Hancock, 2000)

Four Steps for Computation of FSDT Measures

1) Select mapping functions for signal & response dimensions

To assign degrees of (**s**, **r**) membership to events, all possible states of the world and each possible response must be evaluated using a mapping function. For the set **s** a mapping function relates the signal value (**s**) to a variable (or set of variables) that describe the state of the world (see Figure 3). A mapping function for the set **r** relates the response value (**r**) to a response variable (or set of variables). For instance, a mapping function could be based upon confidence ratings of signal presence, a technique used in traditional SDT (Green & Swets, 1966; MacMillan & Creelman, 1991).

2) Assignment of degrees of membership to the four outcomes using mixed implication functions.

H = min (**s**,**r**)
M = max (**s**-**r**, 0)
FA = max (**r**-**s**, 0)
CR = min (1-**s**, 1-**r**)

3) Compute fuzzy Hit, Miss, False Alarm, and Correct Rejection Rates

HR= Σ(Hi)/ Σ(si) for i=1 to N
MR = Σ(Mi)/ Σ(si) for i =1 to N
FAR = Σ(FAi)/ Σ(1-si) for i=1to N
CRR = Σ(CRi)/ Σ(1-si) for i= 1 to N

4) Compute detection theory measures traditional SDT equations.



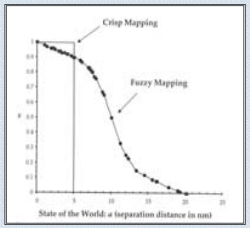


Figure 1: *Sigmoid mapping function relating signal value, s, to each possible value of aircraft separation distance, a, as might occur in air traffic control. The rectangular function represents the standard or crisp SDT mapping of* **s** *to a. Adapted from Parasuraman, Masalonis, & Hancock (2000).*

- An initial experiment investigated the degree to which the FSDT model was sensitive to the classical SDT manipulations of difficulty and response bias. Stimuli employed are shown in Figure 2.
- The same set of data from a discrimination task was analyzed using both FSDT and SDT methods of analysis (see Figure 3a and Figure 3b)



Figure 2. *Stimuli used in an Initial Experiment*

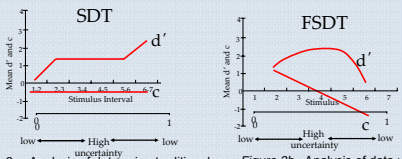


Figure 3a. Analysis of data using traditional signal detection theory

Figure 3b. Analysis of data using fuzzy signal detection theory

Response bias in vigilance has been demonstrated to be influenced by changes in signal probability with time on watch (Parsons, 2001; Vickers & Leary, 1983). Recent experimental work has demonstrated that the FSDT model is sensitive to manipulations of the relative frequency of stimuli that are high or low in threat or 'signalness' (Szalma, Hancock, Mouloua, & Stafford, 2003). Specifically, Szalma et. al. examined the impact of transitions on response bias by varying the frequency distributions for stimulus presentation in a vigilance task. The task required participants to monitor the degree to which two vertically aligned rectangles were similar in color. Observers engaged in two vigils. In the first 20-minute session they monitored a display in which either stimuli high in 'signalness' (i.e., the rectangles were different in color) were more frequent (see Figure 1a) or stimuli low in signal value were more frequent (see Figure 1b). In the second 20-minute session participants either experienced the condition to which they were assigned in the first session or they switched to the other distribution. Observers were not informed of the relative frequency of stimuli or the change between sessions. Details regarding the stimuli and task characteristics can be found in Szalma et al. (2003).

As can be seen in Figure 2, observers who switched from a high to a low frequency of 'signal-like' stimuli became more lenient in the second session than those who experienced the high condition in both sessions. There was a trend for those in the low-to-high transition group to become more conservative in the second session, but this trend was not significant (see Szalma et al., 2003). When the data are analyzed separately for each stimulus, the effect of 'signal' frequency shifts on response bias in both sessions is more pronounced for stimuli that are more 'non-signal-like' (i.e. stimulus 2 and stimulus 3 in Figure 3). This represents an asymmetry in signal frequency effects for stimuli at different ends of the signal dimension. Further research is necessary to identify the factors that control this asymmetry.

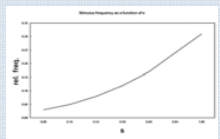


Figure 1a. Relative frequency of presentation as a function of 'signalness' for the 'high-signal' condition

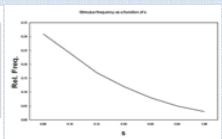


Figure 1b. Relative frequency of presentation as a function of 'signalness' for the 'low-signal' condition

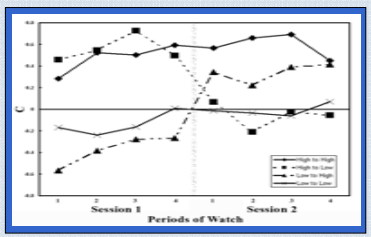


Figure 2. Response bias scores as a function of periods of watch for the four probability transition groups. Data for the two sessions are displayed separately in each panel. Note. High = greater frequency of 'signal-like' stimuli; Low= greater frequency of 'non-signal-like' stimuli (see Figure 1). Each period of watch was 5-min.

Applications of FSDT

Vigilance (e.g. airport luggage screening: what constitutes a weapon?)

Terrorist identification/detection

Preventing potential abuse: Ethnic Profiling. FSDT can be used to determine the degree to which response biases influence terrorist identification decisions.

Threat assessment (e.g. based upon intelligence data)

Combat (e.g. target acquisition; friend/foe identification)

Threat detection/response in police work (see Figure 5)

Example: Tank Identification

Critical to these application will be identifying the mapping functions between the signal dimension and the state-of-the-world in a continuously changing environment in which the definition of a 'signal' changes over time and situations. In other words, the mapping functions may not be stable.



Figure 4. *Tank Image used for identification task*

- When ground troops must identify individuals or objects of relatively unknown hostility, spread over a very large country - confusion from the 'fog of war' can lead to a friendly fire incident.
- Morphing techniques provide an excellent tool for creating a measure of similarity between two distinct objects (e.g. tanks)
- Fuzzy Signal Detection Theory provides a quantitative tool for measuring uncertainty related to the identification of these objects.

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, P.A. Hancock, *Principal Investigator*. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.



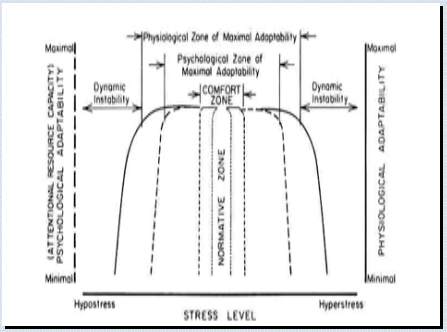
Figure 5. Representation of the range of threat levels facing police officers.

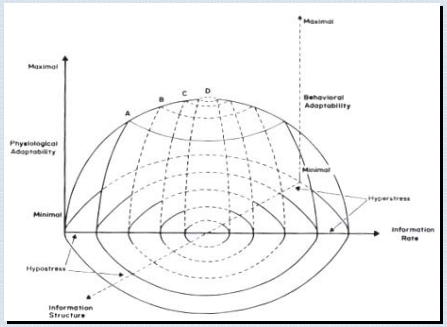
Low Threat/Response — High Threat/Response

Operator Performance Under Stress: A Multiple University Research

P.A. Hancock, J.L. Szalma, T. Oron-Gilad, & the MURI-OPUS Group

Theoretical Foundation
The Stress and Performance Theory
The Hancock and Warm Model





Hancock, P.A., & Warm, J.S., (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.

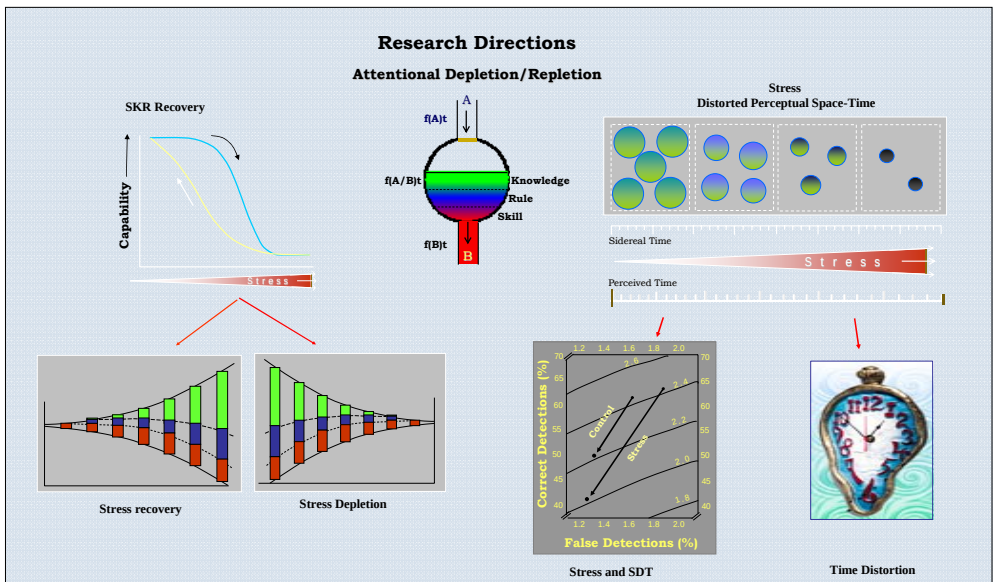
The Hancock and Warm model provides a general , overarching, descriptive approach to characterize the effects of stress on performance. The mediating causal argument is through the depletion of attentional resources. As is evident, there are thresholds indicated by the performance shoulders of the model sometimes identified as the 'extended-U.' It is these threshold of incipient failure that we are now exploring in Laboratory, Simulation, and Field-based research.

The following research goals are currently being addressed:

- Identification of fundamental task characteristics that influence human-machine interaction in task performance under stress
- Increased understanding of how these task characteristics influence attentional narrowing under stress



Research Directions



- To date, this initiative Includes over 30 refereed publications and over 100 conference proceedings and presentation evaluating the impact of stress and workload on performance in both laboratory and field settings.
- Initial tests of hypotheses based on the Hancock and Warm (1989) model have shown support for common spatio-temporal resource depletion. Field work has demonstrated that the precipitous drop in adaptive function occurs under extreme conditions.
- Initial tests of a novel fuzzy signal detection (FSDT) model for assessing performance under stress show support for its utility. Other work has examined the adaptability of teams under stress and the impact of automation reliability on performance. Current efforts and future activities include the time-course of resource depletion and recovery for different forms of information processing. In addition, work on the application of fuzzy signal detection theory to performance assessment will be extended to real-world stimuli. These efforts include laboratory and field settings, as well as information processing tasks required of soldiers.
- A descriptive framework for the relation between sources of stress, information processing requirements, and moderators *of that relation (see The Matrix section for more details).*

The Matrix

Detailing the multidimensional nature of stress and performance, and predicting how performance is debilitated or facilitated by stress can be arduous task. This task becomes more and more difficult when considering the potential interactions between environmental, task-based and social stressors and the information processing requirements of tasks sub-serving perception, cognition and action. The complexity of performance prediction increases exponentially when one includes factors that moderate the stress-performance relationship, and permit performance effectiveness to be maintained, albeit at the potential cost of processing efficiency.

The majority of research to date has primarily focused upon the effects of a single environmental stressor on 'isolated' information processes (Hockey & Hamilton, 1983). Moreover, the effects of task-based stress have seldom been considered (Hancock & Warm, 1989). While limited information is available on the effects of environmental stress such as heat, toxic substances, noise, and vibration much of the research has attempted to elicit threshold limits as opposed to determining the degree of degradation per unit of time making performance prediction difficult (for a review, see Pew & Mavor, 1998).

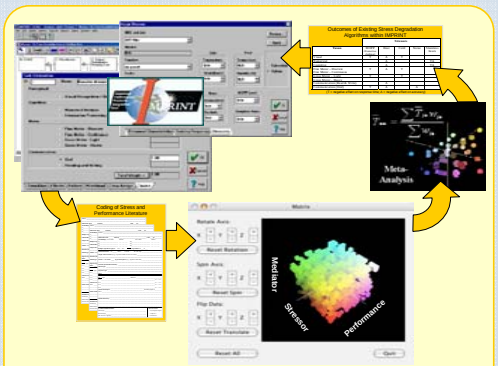


Figure 1: Evidence-based extension to IMPRINT

- While stress can both facilitate and debilitate performance, the negative effect upon memory recall demonstrated by the group and grand weighted mean effect sizes is consistent with the expected direction of effect (i.e., Hockey & Hamilton, 1983). Previously, noise has been shown to increase the use of order information while decreasing semantic organization, and is typically associated with poor short term recall (Broadbent, 1981). Hockey and Hamilton's (1983) negative assessment of noise upon performance, however was based upon the presentation of continuous noise ranging from 80 to 100dB. Our meta-analysis data extend these research findings by examining the magnitude of effect on memory recall as noise intensity increases.
- The data indicate that the size of effect is related to the magnitude of the difference between noise intensity conditions. For instance the size of effect, and the decrement in performance, when moving from 85-95dB to >95dB is almost twice as large as when moving from <55dB to 85-95dB. What is more, both contrasts were significantly more detrimental to recall than when moving from <55dB to 66-78dB.

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, P.A. Hancock, *Principal Investigator*. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

Fuzzy Signal Detection Theory: Tests of Assumptions and Comparison to Traditional Analyses

Lauren L. Murphy James L. Szalma Peter A. Hancock

Both Traditional Signal Detection Theory (SDT) and Fuzzy Signal Detection Theory (FSDT; see Parasuraman, Masalonis, and Hancock, 2000) methods of analyses were compared using a duration discrimination task of brief light flashes. Both difficulty and response bias (by instruction) were manipulated. The main research question for FSDT was whether the assumptions of traditional SDT hold for this new model. The current study indicates that at both difficulty levels and for all six participants, the assumption of normally distributed noise and signal plus noise distribution holds. In addition, the equal variance assumption holds for the more difficult discrimination for all participants. Results for the traditional SDT analysis varied across subjects, and may depend upon how the stimulus dimension is categorized for treatment using traditional methods. These results suggest that traditional SDT may have limited utility in cases where stimuli can vary in their signal value.

Fuzzy Signal Detection Theory: Tests of Assumptions and Comparison of Traditional Analyses

L.L. Murphy, J.L. Szalma, & P.A. Hancock

Abstract

Both Traditional and Fuzzy Signal Detection Theory (SDT) methods of analyses were compared using a duration discrimination task of brief light flashes. Both difficulty and response bias (by instruction) were manipulated. The two methods clearly provided different estimates of SDT parameters. The Fuzzy SDT analyses inflated performance while the traditional SDT analyses underestimated performance. In some cases performance on the 'easier' condition was worse than on the more 'difficult' condition. This may be due to differences in the range of durations discriminated, which was narrower in the 'easier' condition. Range of stimuli used can impact performance, although it or maybe an artifact of the procedure for the crisp analysis. Fuzzy SDT met the assumptions of traditional SDT, the assumption of normally distributed noise and signal plus noise distribution holds. In addition, the equal variance assumption holds for the more difficult discrimination for all participants.

Introduction

Similar to Signal Detection Theory (SDT), Fuzzy Signal Detection Theory (FSDT) provides a useful measurement tool for evaluating human and machine performance in both simple and complex systems. However, unlike SDT, FSDT is not limited to forcing the state of the world into mutually exclusive categories (i.e., signal versus non-signal; threat versus non threat; etc.).

SDT does not capture the uncertainty inherent in real world stimuli. Whereas, FSDT, through the representation of the stimulus dimension as continuous rather than categorical, captures the uncertainty inherent in real-world stimuli.

FSDT combines elements of SDT with those of Fuzzy Set Theory, thus category membership is not considered mutually exclusive and stimuli can therefore be simultaneously assigned to more than one category thereby capturing this uncertainty (Parasuraman, Masalonis, and Hancock 2000; see also Hancock, Masalonis, & Parasuraman, 2000).

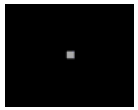
In FSDT, a stimulus selected from a range of stimuli may be categorized as both a correct detection and a false alarm depending on the degree to which the stimulus represents a critical event. For instance, a convenient range for a stimulus dimension is one in which the strength of the stimulus varies from 0 (a definite non-signal) and 1 (a definite signal), with a signal value of .5 representing maximal uncertainty in the stimulus itself. That is, a stimulus with a signal value of .5 has properties of both a non-signal and a signal to an agreed degree. Implicit in this model is the assumption that signal uncertainty exists not only within the observer but in the stimulus dimension itself.

This experiment represents a test of FSDT using a perceptual signal detection task. In this study fuzzy stimulus and response dimensions were established (seven categories for each), and difficulty (discrimination) and bias (instructional set) were manipulated to investigate the degree to which the FSDT model is sensitive to these classical SDT manipulations. The FSDT analysis was compared to an analysis using traditional SDT, using methods outlined in MacMillan and Creelman (1991).

Method

Participants

Six students from the University of Central Florida (3 men and 3 women, mean age = 19 years) participated in this study for monetary compensation.



Stimuli

The stimuli employed in this experiment consisted of a 8x8 cm light gray square with a black surround. The square was presented at seven different durations depending on the difficulty level. Participants were not informed of the number of stimulus categories, only that the stimuli would vary between two extremes in duration.

For each condition the event rate was 21 event/min, and each of the 7 different stimuli were presented 120 times during each session, totaling 840 trials per condition. The order in which the stimuli were presented within a condition was randomized.

Procedure

The first day, participants received instructions and completed two 40-min practice trials of the task. Both difficulty levels of the task (less difficult vs. more difficult) were presented in the practice session, the order in which the participants received the difficulty levels was counterbalanced.

In the remaining three days of the experiment, participants engaged in a total of six 40-min detection tasks with 5 min breaks after every ten min consisting of two temporal difficulty levels (less difficult, stimulus presented at 7 durations ranging from 200 ms to 680 ms with a delta of 80 ms) vs. more difficult (stimulus presented at 7 durations ranging from 200 ms to 320 ms with a delta of 20 ms) and three levels of instruction bias (lenient, conservative and neutral).

Participants received the treatments over 3 days during a two week period. Each day they received one of the bias manipulations and both difficulty levels. The order in which the participants received the different treatments was counterbalanced using a Latin Square design.

Task

The task required the participants to monitor the relative duration that the square of light flashed on the screen. Participants were instructed to respond to stimuli by rating the degree to which each stimulus was a non-signal (appearing for a short period) versus a signal (appearing for a long period) by pressing keys 1 through 7, with the response '7' indicating that the stimulus was definitely a signal.

Bias by Instruction

For the manipulation of response bias, three instructional sets were employed to induce lenient, unbiased, and conservative responding.

During the task with the ***lenient*** instruction set participants were informed that they would receive (10) points for each correct identification, which meant that they correctly identify the duration of the flash of light. However, they were told that they would loose (-10) points for each missed signal, which meant that they underestimated the duration to which the flash of light appeared. Finally, they were informed that they would be penalized (-1) point for each false alarm, which meant they overestimated the duration to which the flash of light appeared.

Payoff Matrix for a 2x3 Repeated Measures Design

Bias by Instruction

Difficulty	Conservative	Neutral	Lenient
High	10 Hit -1 Miss -10 FA	1 Hit -1 Miss -1 FA	10 Hit -10 Miss -1 FA
Low	10 Hit -1 Miss -10 FA	1 Hit -1 Miss -1 FA	10 Hit -10 Miss -1 FA

During the ***neutral*** (unbiased) condition participants were informed that they would receive (1) point for each correct identification. However, they were told that they would be penalized (-1) point for each missed signal and that they would be penalized a (-1) point for each false alarm.

During the ***conservative*** condition participants were informed that they would receive (10) points for each correct identification. However, they were told that they would be penalized (-1) point for each missed signal and (-10) points for each false alarm.

Results

Currently, efforts are underway to develop a statistical program that can test the significance of the fuzzy SDT ROC curves quantitatively (Wickens, 2002). Due to the constraint of appropriate estimation procedures, the data analyzed using fuzzy and traditional signal detection methods will be compared qualitatively using FitRoc: Parameter Estimation for Gaussian Signal Detection Model (Wickens, 2002), version 1 program which provides (1) a test of how well an ROC curve fits the Gaussian model ($\chi 2$), (2) an estimate of perceptual sensitivity (Az), and (3) the intercept and slope for the z-score form of the ROC are estimated by the program (zH = a + b zF). Tables 1 and 2 report goodness of fit, intercept (e.i., 'a'), slope (e.i., 'b'), and perceptual sensitivity for the traditional SDT method for the easier and more difficult conditions. Similarly, Tables 3 and 4 report these statistics for the fuzzy SDT method of analyses.

Formulas for calculating sensitivity and response bias were the same for both fuzzy SDT and traditional SDT analyses. The two models differ in how hit and false alarm rates are calculated. The formulas for calculating FSDT measures were obtained from Parasuraman et al. (2000), and the method for computing hit and false alarm rates followed procedures outlined by MacMillan and Creelman (1991).

For traditional SDT, results indicated for most subjects (except subject 2) perceptual sensitivity, A(z), was higher in the difficult condition compared to the easy condition, refer to Tables 1 and 2. Similarly, this was the case for fuzzy SDT, subjects (except subject 2) appeared to perform better in the difficult condition than the easy condition, refer to Tables 3 and 4. Additionally, perceptual sensitivity was higher for all subjects for fuzzy SDT compared to traditional SDT for both difficulty levels.

Table 1. *Traditional SDT method for the easy condition*

Subject	χ^2	A(z)	a	b
1	0.00[u]	0.661	0.449	0.668
2	25.95[u]	0.748	0.790	0.629
3	14.28[u]	0.546	0.135	0.607
4	55.08[u]	0.800	2.053	2.224
5	0.59[u]	0.615	0.362	0.728
6	19.78[u]	0.659	0.534	0.839

Table 2. *Traditional SDT method for the difficult condition*

Subject	χ^2	A(z)	a	b
1	17.74[u]	0.707	0.799	1.129
2	0.23[u]	0.698	0.854	1.310
3	1.24[e]	0.599	0.353	1.00
4	1.32[e]	0.681	0.666	1.00
5	7.10[e]	0.616	0.416	1.00
6	2.78[e]	0.631	0.474	1.00

Table 3. *Fuzzy SDT method for the easy condition*

Subject	χ^2	A(z)	a	b
1	0.96[u]	0.808	0.948	0.435
2	6.96[u]	0.837	1.058	0.398
3	2.17[u]	0.793	0.905	0.481
4	60.63[h]	0.785	1.115	1.00
5	0.02[u]	0.842	0.248	0.742
6	3.54[u]	0.815	1.034	0.576

Table 4. *Fuzzy SDT method for the difficult condition*

Subject	χ^2	A(z)	a	b
1	2.60[e]	0.819	1.290	1.00
2	1.18[e]	0.823	1.309	1.00
3	1.96[e]	0.793	1.156	1.00
4	0.61[e]	0.805	1.218	1.00
5	0.98[e]	0.796	1.172	1.00
6	0.02[e]	0.687	0.541	0.47

As shown in Tables 1, 2, 3, and 4, goodness of fit intercept and slope, revealed that for all subjects (except participant 4 in the easy traditional condition) and for all conditions, the data fit the assumption of normality; the noise and signal plus noise distributions were normally distributed. Interestingly, the Tables reveal that the data fits the equal variance assumption, see c2, for all six subjects, but only for the difficult condition for both traditional and fuzzy SDT.

Discussion

The traditional and fuzzy methods clearly provide different estimates of SDT parameters. The inflation of performance with the fuzzy SDT model is in accordance with prior findings (Murphy, Szalma, Hancock, 2003). It may be due to allowing a degree of correct detection rather than a miss, and similarly allowing only a degree of false alarm rather than a 'full false alarm' (c.f., Masalonis & Parasuraman, 2003) That is, partial hit rates in fuzzy SDT are categorized as whether full hits or full misses in the crisp analysis, and partial false alarms are categorized in the traditional SDT analyses as full false alarms or correct rejections. Thus, the traditional SDT analyses of this kind of data may underestimate the performance of observers.

The different results across observers for the difficulty manipulation using the crisp analysis was unexpected. In two cases performance on the 'easier' condition was worse than on the more 'difficult' condition. This may be due to differences in the range of durations discriminated, which was narrower in the 'easier' condition. Prior research has shown that the range of stimuli used can impact performance (Braida and Durlach, 1972). However, this does not explain why such results were obtained with only two of the six observers. It is possible that it is merely an artifact of the procedure for the crisp analysis. The possibility for range effects will be explored in future experimentation.

A central question for fuzzy SDT is whether the assumptions of traditional SDT hold for this new model. The current study indicates that at both difficulty levels and for all participants (except the easy condition, participant 4), the assumption of normally distributed noise and signal plus noise distribution holds (see c2 tests in Tables 3 and 4). In addition, the equal variance assumption holds for the more difficult discrimination for all six participants. Future studies will examine the effects of the size of both the stimulus and response sets on signal detection using fuzzy SDT.1

Acknowledgement

This work was facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, Dr. P.A. Hancock, *Principal Investigator*. The assistance of the Department of Psychology at University of Central Florida is also gratefully acknowledged. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

References

Braida, C.D., & Durlach, N.A. (1972). Intensity perception. II. Resolution in one-interval paradigms. *Journal of Acoustical Society of America*, **51**, 483-502.

Hancock, P.A., Masalonis, A.J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science*, **1**, 207-230.

Macmillan, N.A., & Creelman, C.D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Murphy, L.L, Szalma, J.L, & Hancock, P.A. (2003). Comparison of Fuzzy Signal Detection and Traditional Signal Detection Theory: Approaches to Performance Measurement. *Proceedings of the human factors ergonomics society 47th annual meeting: Denver, CO*, 1967-1971.

Myers, J.L., & Well, A.D. (1995). Research Design and Statistical Analysis. (2nd Ed.). New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, **42**, 636-659.

Wickens, T.D. (2002). *Elementary Signal Detection Theory*. Oxford University Press.0

Psychological Strategies for the Defence Against Terrorism

Mark E. Koltko-Rivera, Ph.D.
Director of Research
Professional Services Group, Inc.
P.O. Box 4914
Winter Park, Florida 32793-4914 USA

P. A. Hancock, D.Sc., Ph.D.
Department of Psychology and Institute for Simulation and Training
University of Central Florida
P.O. Box 161390
Orlando, Florida 32816-1390 USA

E-mail: mark@professionalservicesgroup.net / phancock@pegasus.cc.ucf.edu

SUMMARY

Various subdisciplines of psychology are relevant to the defence against terrorism, in terms of anti-terrorism, counter-terrorism, and terrorism consequence management. Anti-Terrorism: Psychological methods can be applied to reduce vulnerabilities to attack and to encourage the general public to identify infrastructure and other vulnerabilities. Counter-Terrorism: Psychological techniques are available to assess and improve terrorism awareness in the general population. The detection performance of counter-terrorism personnel can be improved: psychological methods can enhance situation awareness, situated cognition, detection capabilities, and decision-making; automated expert system tools employing fuzzy signal detection can assist personnel; other psychological techniques can enhance individual and team function, personnel selection and training. Psychological principles can also be applied to obstruct and impede terrorist functioning. Consequence Management: Psychological methods can be used to enhance capabilities of first responders, improve escape and evacuation procedures for civilians, promote resilience in the general population, and treat victims of terrorism more effectively. We propose possible configurations for psychological consulting teams who would help defence authorities use these strategies to address terrorist activity.

1.0 INTRODUCTION

Defence against terrorism (DAT) is often discussed in terms of technological innovations and technical barriers to terrorist success. This includes considerations of surveillance and intelligence; naval, air, space, and land-based military sciences; information and communication technology; security engineering; radiation technology; sensors and telemetry; imaging technology; weapons science, and so forth. As valuable as these disciplines are for DAT, it is our opinion that the traditional social sciences have critical contributions to make in this area. In particular, we propose that psychology has much to offer in furthering DAT. As we hope to demonstrate, psychological comprehension is crucial for successful DAT in all its forms.

Paper presented at the RTO SCI Symposium on "Systems, Concepts and Integration (SCI) Methods and Technologies for Defence Against Terrorism," held in London, United Kingdom, 25-27 October 2004, and published in RTO-MP-SCI-158

Much of what has been written about psychological approaches to terrorism has focused on understanding the internal psychology of the terrorist, the sociocultural roots of terrorism, the social psychology of terrorism's aftermath, and public policy about terrorism [1-11]. As valuable as these contributions are, our intent here is to embrace a much broader view of psychology's potential impact on DAT. In this limited space, we cannot deal with all of psychology; thus, we 'bracket,' or leave for other venues, the pressing questions of how terrorism begins and how its psychosocial origins can be addressed. Rather, we focus here on how psychology can help security and intelligence personnel prevent or respond to specific terrorist acts, and how psychology can help ameliorate the consequences of those acts.

Consider the stereotypical psychologist: a clinician who reclines in a padded armchair, analyzing societal problems in terms of difficulties with early toilet training, and reacting to every client statement with such canned responses as "How does that make you feel?" This stereotype not only seriously misrepresents even clinicians, it ignores the great number of experimental psychologists and related neuroscientists who do not engage in clinical work at all. Psychology is an extraordinarily wide umbrella covering many subdisciplines, such as human factors and ergonomics (i.e., the science of human performance and human-machine interaction), cognitive psychology, neuroscience, artificial intelligence, social and personality psychology, consumer psychology, counseling and psychotherapy, and cultural differences in behavior, to mention only a few subdisciplines. Psychologists design measures of attitudes and behavior, study the interaction of humans with machines, and design research to test programs to improve human performance in many kinds of environments. As we demonstrate below, many subdisciplines within psychology are applicable to DAT.

In this work, we outline contributions that the discipline of psychology can make within each of the three facets of DAT: Anti-Terrorism, Counter-Terrorism, and Consequence Management. Throughout our presentation, we identify psychological methods, specialty areas, and topics (italicized below) that are particularly pertinent to DAT. It is our hope that civil and military authorities will consider implementing psychological approaches and technologies, as these authorities grapple with the challenges of defending against terrorism.

A few words are necessary concerning our approach in this work. Being mindful that a paper such as this may find its way into the hands of many different kinds of DAT personnel, we have chosen to reference this work rather thoroughly, including not only basic reference works, but also specialized texts, and recent conference proceedings. Our intent in doing this is to enable the reader, who may need to develop a brief on potential psychological contributions for others in the chain of command, to self-educate in the areas we touch upon, should the need arise. In addition, we have defined as "psychological" much research and theory where psychology overlaps with other fields. (For example, although we describe artificial intelligence, or AI, as a psychological specialty area, AI obviously owes a great deal to the discipline of mathematics as well). In doing this, our intention is not to deemphasize the importance of other fields, but to be inclusive in demonstrating the contributions that psychology has to make to DAT.

2.0 PSYCHOLOGY AND ANTI-TERRORISM

Since the founding of NATO, its member nations have been open societies, with relatively permeable borders, unsecured infrastructures, easily accessed information and communication systems, and populations who have enjoyed life without open warfare within their own borders for now more than five decades. The threat of terrorism exposes the vulnerability of open, peaceful societies in multiple ways. Unsecured infrastructures and easily accessed information and communication systems are each eminently vulnerable to attack. Easily traversed borders facilitate travel for friend and foe alike. Populations that are accustomed to years of

domestic peace may be unfamiliar with, or insensitive to, signs of terrorist preparation.

Anti-terrorism involves defensive measures to reduce the vulnerability to attack presented by populations, territory, infrastructure, and information and communications systems. Psychological methods can be used to promote anti-terrorism by furthering two objectives: reducing and identifying vulnerabilities.

2.1 Apply Human Factors and Ergonomics Principles to Reduce Vulnerabilities

Post-9/11, there has been much effort devoted to designing or renovating buildings to address security issues; such concerns are evident in recent design criteria for U.S. federal buildings [12]. The disciplines of psychology and engineering overlap in the subdiscipline of *human factors and ergonomics (HF/E)*. This specialty uses knowledge of human abilities and limitations for the design of systems, organizations, jobs, machines, and other physical objects for safe, efficient, and comfortable human use (adapted from [13], p. 4). One sub-specialty within this broad area involves the use of HF/E principles in architecture and interior design; these principles can be used to implement security and evacuation enhancements as part of a complete design strategy [14]. This can be seen as a variation of *environmental design research* in psychology, which focuses on the quality of fabricated environments [15]. Principles from HF/E and environmental design research should be implemented, both to design new structures and systems, and to retrofit older structures and systems, to reduce vulnerability to terrorism. For example, a wealth of suggestions have been made to increase aviation security through the implementation of HF/E principles [16].

One type of anti-terrorist activity involves the use of so-called "red teams." These are teams of anti-terrorist personnel devoted to discovering and correcting a system's areas of vulnerability, before terrorists find these liabilities. Although often applied against cyberterrorism, red teams can in principle be applied against any type of terrorist threat. HF/E researchers have studied anti-cyberterrorism red teams, an activity that has yielded suggestions for improved red team effectiveness [17]. Similar research studies would likely improve the function of red teams deployed against other types of terrorism, as well.

2.2 Enlist the Aid of the Populace in Identifying Vulnerabilities

As the terrorists themselves have shown, it is possible for low-tech terrorism strategies to circumvent relatively sophisticated security procedures. It would be appropriate for security authorities to attempt to anticipate such attacks, by enlisting the combined intellectual power of their own populace through the use of *behavioral reinforcement techniques* [18, 19]. That is, through the use of mechanisms of reward and schedules of reinforcement, security authorities may encourage members of the public to report potential vulnerabilities in infrastructure, target-likely buildings, and so forth. (This idea bears a distant familial resemblance to various "Crimestoppers" and "Crime Tips" programs, in which members of the public may anonymously report information bearing on crimes, for monetary reward; such programs have resulted in many prosecutions. Our idea, however, involves the proactive detection of vulnerabilities, rather than the reactive reporting of criminal activities.) A variation on this idea would be to use reinforcement techniques to shape the behavior of volunteer security forces patrolling unsecured infrastructure. This is the idea behind the Airport Rangers, a volunteer force on horseback that patrols an international airport in Texas [20]. Another variation on this idea would be to encourage selected professionals (e.g., human factors professionals, structural engineers, computer professionals) to come forth with their observations of systemic or particular vulnerabilities [21]— volunteer professional red teams, as it were.

Some have written of the "impossibility of completeness" inherent in attempts to prevent terrorism at or along the way to potential targets ([10], p. 38). Although absolute completeness may be impossible, enlisting the

assistance of a national or local population in identifying problem areas will surely yield valuable information about infrastructure vulnerabilities that we would not otherwise possess.

3.0 PSYCHOLOGY AND COUNTER-TERRORISM

Each terrorist activity has a cyclical form, involving the stages of planning, preparation, execution, escape, and evaluation. (Of course, suicide attacks do not involve escape or evaluation by the terrorists most directly involved.) Successful completion of the cycle means that the terrorist can begin the cycle anew, perhaps spiraling to some larger scale of attack. However, each stage presents at least some opportunity for security authorities to interfere with the cycle. Interrupting this cycle comprises counter-terrorism.

More precisely, counter-terrorism involves offensive measures to track down, prevent, deter, and interdict terrorist activities. Psychological methods can be applied to counter-terrorism to meet three important objectives: increased population alertness to terrorist activity, improved detection performance of security and intelligence personnel, and increased obstruction of terrorist functioning.

3.1 Make the Populace Alert to Signs of Incipient Terrorist Activity

Lightning rarely strikes from a clear sky. Similarly, terrorism rarely if ever occurs without preparation. Terrorists travel. Materials, transportation, lodging, training, and access to targets must be purchased. Targets must be researched and surveilled. Destructive materials must be appropriately placed. An alert populace may be able to detect and report preparations for terrorist activity before terrorist acts are executed. This may involve reporting specific patterns of activity (e.g., unusual purchases). In addition, an alert populace may be able to detect and report events that indicate that a terrorist activity is in the midst of execution. For example, this may involve reporting the presence of suspicious objects (e.g., abandoned packages). Several psychological methods and specialty areas are applicable to the task of making a population more alert to signs of terrorist preparations, through the assessment and improvement of terrorism awareness.

3.1.1 Assess Terrorism Awareness

Before we can change behavior towards an ideal, we must have a sense of what the current state is, and how this state deviates from the ideal we have established. *Assessment techniques* can be devised to gauge levels of individual and group awareness regarding signs of potential terrorism. The psychological subdiscipline *psychometrics* possesses well-established techniques to develop assessment methods that are valid and reliable [22, 23]. The basic objectives of assessment research should be (1) to define what is meant by terrorism awareness, both in general and in terms of its components; (2) to develop valid and reliable assessment techniques for each of the components of terrorism awareness; (3) to gather valid data about the populace's terrorism awareness using these techniques; and, (4) to identify strengths and weaknesses in the populace's terrorism awareness. Once these basic objectives are met, a further worthy objective would be (5) to identify conditions that tend towards greater and lesser terrorism awareness.

3.1.2 Improve Terrorism Awareness in the General Population

Once we have a sense of the level of terrorism awareness present in the populace, and a sense of what components of terrorism awareness show particular need for improvement, we are in a position to attempt to improve the social level of terrorism awareness. Three psychological specialty areas are relevant to this area: *educational psychology, methods of persuasion,* and *program evaluation*.

One approach to improving terrorism awareness involves overt instruction ([10], p. 39). Psychology has a long history of research into learning methods as an aspect of *educational psychology*. Much of this research addresses ways to enhance or optimize learning [24, 25], and this research should be applied to designing instructional methods for increasing the terrorism awareness of the general public. Although much educational research deals with schoolchildren, a great deal of this research is applicable to adults as well. (For that matter, classroom instruction of children in terrorism awareness should not be overlooked.)

Another approach to improving the public's counter-terrorism behavior involves *methods of persuasion*, a topic of psychological research for a long time [26-31]. A subfield of *consumer psychology* that combines the psychology of instruction with that of persuasion is the area of *"social marketing,"* which has the objective of changing personal and social behavior, typically to further some societal goal [32-36].

Of course, a program to improve terrorism awareness is only as good as the changes it actually creates in personal and social behavior. To assess these changes, *program evaluation* methodologies are crucial [37-39].

3.2 Improve Detection Performance of Counter-Terrorism Personnel

Psychology has much to offer in terms of improving the performance of counter-terrorism personnel, who are in the front lines of terrorism prevention. This applies both to personnel in immediate security, such as airport baggage screeners, and to more meta-level professionals, such as intelligence agents. Psychology's potential contributions involve enhancing the psychological capacities of these personnel for detection, giving these personnel improved tools to use for detection, helping these personnel to perform better as parts of teams, helping administrators to select appropriate personnel for counter-terrorism, and training personnel more effectively following their screened selection.

3.2.1 Enhance Detection Capabilities of Counter-Terrorism Personnel

It is well-known that baggage screeners' detection of weapons carried by passengers at American airports was not optimal, even as late as 2003 [40]. It is also well-known that American intelligence agencies' detection of the threat posed by terrorists in 2001 was faulted in the 9/11 Commission's report [41]. These two portions of the counter-terrorism spectrum share some characteristics. At a high level of abstraction, all counter-terrorism personnel are engaged in the detection of relatively rare but potentially catastrophic events. For baggage screeners, the task is to detect the presence of small weapons of interpersonal violence or mass destruction; for intelligence agents, the task is to detect patterns of activity suggestive of preparations for terrorist activity. Ultimately, then, the task facing each type of personnel involves the use of two crucial psychological faculties involved in detection of a terrorist threat: attention (that is, the capacity to detect and attend to crucial stimuli) and decision-making. *HF/E* and *cognitive psychology* are subdisciplines that address these crucial psychological capacities.

HF/E has amassed considerable research-based insight that can be used to address several issues relevant to the attentional performance of counter-terrorism personnel. These issues include vigilance [1, 16, 42-44], focused attention [45], selective and divided attention [46], stress and fatigue [47, 48], and signal detection [49]. HF/E researchers have investigated very specific aspects of cognition and human performance relevant to counter-terrorism; for example, some researchers have investigated the function of mental rotation within a baggage screening task [50].

Some HF/E scientists are involved in research regarding *neuroergonomics*, the study of the brain and behavior when a person is working [51, 52]. This research is relevant to counter-terrorism. For example, some current

neuroergonomic research involves the study of brain function during baggage screening tasks, as performance degrades with fatigue [53]. More generally, the study of continuous wakefulness and consequent fatigue, and the effect of these variables on performance of military personnel, is a long-standing area of concern in HF/E [54-56]. This research is applicable to the improvement of performance for counter-terrorism personnel.

There are different ways in which HF/E researchers have addressed attention. One involves the concept of *situation awareness*:

> The formal definition of situation awareness is 'the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.' The term *situation awareness* comes from the world of the military pilot, where achieving high levels of situation awareness was found to be both critical and challenging early in aviation history.… The importance of situation awareness as a foundation for decision making and performance applies to almost every field of endeavor. ([57], p. 13, citation and some italics omitted, abbreviation expanded)

A great deal of HF/E research and theory has focused on situation awareness ([57-64]; see also multiple papers in [65, 66]). It is likely that some of the research that has focused on situation awareness in military personnel [67-70] will prove useful in improving the situation awareness of counter-terrorism personnel.

Another way in which HF/E scientists have addressed attention recently is through the concept of *situated cognition* [71, 72]. At the risk of oversimplifying, we may say that if situation awareness is a still photograph, situated cognition is a motion picture. Situated cognition follows the way in which information about a given situation is processed, from "ground truth," through several conceptual lenses, to a given decision maker's projection or interpretation of the situation. It is likely that research involving situated awareness will be useful in improving the performance of a wide variety of counter-terrorism personnel.

HF/E researchers have also emphasized research on decision-making ([73, 74]; see also multiple papers in [65, 66]). In particular, human factors researchers have focused much attention on decision-making under stressful conditions, often in military contexts [75-79]. Much of this research is relevant to the decision-making challenges facing counter-terrorism personnel.

Some HF/E research has specifically focused on describing and improving the work of intelligence analysts through focusing on cognitive factors in intelligence analysis [80-86]. *Cognitive modeling* has been applied successfully to understanding and improving the function of, for example, aviation crews [87], and is beginning to be applied to intelligence analysts [88]. It is to be expected that cognitive modeling research will improve performance when applied to counter-terrorism teams and individual personnel.

As it happens, situation awareness and situated cognition may be considered as special cases of more general constructs, such as directed consciousness [89, 90]; this means that a great deal of cognitive research is relevant to the task of improving detection capacity in counter-terrorism. *Cognitive psychology* researchers have amassed an enormous literature of laboratory studies concerning attention [91-94], decision-making [95-98], and reasoning [99-104]. Application of this research to counter-terrorism may result in improved detection performance for security personnel and intelligence agents.

3.2.2 Equip Counter-Terrorist Personnel with Psychologically Sophisticated Automated Tools

As mentioned earlier, to speak highly abstractly, counter-terrorism personnel at all levels are involved in

signal detection. However, especially for intelligence agents, the task involved departs significantly from the classical signal detection paradigm. It is not just that there is a great deal of noise obscuring the signals of preparation for terrorist activity. To make the situation much more complicated, the signals themselves are not necessarily signals, as it were. To use a simple example, sometimes a person purchasing a large amount of agricultural fertilizer simply has a large farm; sometimes such a person is using the fertilizer to produce powerful explosives for terrorist activity. The purchase of large quantities of fertilizer, then, is not a straightforward signal of terrorist activity. Rather, such a purchase is what is technically referred to as a "fuzzy signal": sometimes the signal denotes terrorism, but most often it does not.

To supplement classical signal detection approaches, HF/E researchers have been working to develop *fuzzy signal detection* theory and applications [105, 106]. This approach can be implemented on *artificial intelligence* platforms derived from *cognitive psychology* [107] to create *expert systems* for detection of fuzzy signals [108]. Thus, it would be possible to develop an automated expert system to analyze data (e.g., patterns of purchases, travel, Internet activity) with fuzzy signal detection techniques, and alert counter-terrorism authorities to the detection of patterns that suggest preparations for terrorist activities. (Elsewhere in this symposium, the first author describes the conceptual outlines of such a system; [109]). Use of expert systems implementing fuzzy signal detection also has been advocated for commercial aviation security [42].

3.2.3 Enhance Team Function of Counter-Terrorism Personnel

Counter-terrorist personnel typically function as members of teams. Complicating the issue, members of a given team (e.g., intelligence analysts) may be scattered across several physical locations, and thus must function as distributed or virtual teams. In addition, some teams include "members" who are artificial intelligence constructs, or *(artificially) intelligent agents*. Despite these complications, effective team performance is essential to all counter-terrorism functions.

Researchers in *HF/E* and *industrial-organizational* (I/O) psychology have studied *team functioning* for some time, with an eye toward assessing and improving team function ([16, 110-115]; see also multiple papers in [64, 65]). Some research has focused specifically on distributed or virtual teams, and teams that include (artificially) intelligent agents [116-120]. A great deal of this research is applicable to the challenges faced by counter-terrorist teams.

As counter-terrorism teams increasingly become multicultural or multinational in composition, issues of cultural *worldview* and *acculturation* can be expected to affect team performance. Worldviews are sets of assumptions about physical and social reality that have important effects on behavior [121]. Research demonstrates that some dimensions of worldview affect team performance [122, 123]. It is expected that future research will show that team performance is affected by several more dimensions of worldview, and by differences within a team in terms of worldview and acculturation [124]. Use of psychological techniques to address worldview and acculturation may enhance counter-terrorism team performance.

3.2.4 Use Psychologically Sophisticated Methods to Select and Train Counter-Terrorism Personnel

Present indications suggest that the need for all types of counter-terrorism personnel will continue to increase. counter-terrorism work includes some very stressful occupations; for all counter-terrorism work, the consequences of performing well as opposed to poorly may determine the difference between life and death, perhaps for thousands of individuals. It is thus worthwhile to put a great deal of thought and effort into appropriate personnel selection, training, and development, for all levels of counter-terrorism personnel.

Personnel psychology has amassed a large research base, and is an area of very active research with regard to recruitment and selection [1, 125, 126]; some of this research focuses on military contexts [127]. Recent years have seen a great deal of sophisticated research involving the application of *personality psychology* to personnel selection [128, 129], as well as continued research into the relationship of various personal characteristics to performance; much of this research has occurred in military contexts [130-133]. A great deal of this research is directly applicable to the matter of selecting new counter-terrorism personnel.

Training and development also has long been a concern of psychological research, both for I/O psychologists generally [134], and for HF/E researchers seeking to improve military training [135]. Of course, the research mentioned above addressing ways to enhance learning and instruction [24, 25] should be applied to the instruction of counter-terrorism personnel. Lately, some research activity has focused on evaluating the effectiveness of training, specifically in military contexts [136]. Here, too, much accumulated research is directly applicable to training of counter-terrorist personnel.

Simulations, including virtual role playing, can be a useful component of a training program [137]. In recent years, HF/E researchers have used the *scaled worlds* approach to designing computerized simulations for training, often in military settings [138]. Scaled world simulations have also been developed specifically for first-responder functions within a terrorism context [139]. We expect that further use of this technology will be effective in training all levels of counter-terrorism personnel.

The use of *virtual reality* (VR) for training has been the subject of a great deal of research [140], including some research specifically focused on VR training for high-risk jobs [1], for work involving teams [141], and for national defence [142]. We expect that VR training will prove to be effective in training counter-terrorism personnel.

One aspect of training and development that would be useful for counter-terrorism personnel involves *cognitive enhancement*. This is our term for several lines of scientific research that focus on the nature or enhancement of personal intellect or intelligence [143-145], creativity [145-149], intuition [150], and even wisdom [145, 151-152]. Each of these cognitive capacities can be used to detect or even anticipate possible terrorist activities.

A creative, imaginative approach to anticipating potential terrorist activities is in order. As the 9/11 Commission Report noted, “Imagination is not a gift usually associated with bureaucracies. … [However,] it is … crucial to find a way of routinizing, even bureaucratizing, the exercise of imagination” [41, p. 344]. Such exercises could prove highly useful in anticipating potential terrorist activities. For example, occasionally it has been claimed that no one could have predicted that the 9/11 terrorists would use airplanes as missiles to attack the World Trade Center. In fact, however, the use of hijacked passenger jet airplanes as missiles in an attack on the World Trade Center was anticipated over six months before the 9/11 attacks—in a television drama. The creativity that the television writer and producer Chris Carter showed in the pilot episode of his short-lived television series, “The Lone Gunmen” [153], would be useful for counter-terrorism work.

3.3 Obstruct Terrorist Functioning

Broadly speaking, there are at least two ways in which psychological research and theory may be applied to the task of obstructing terrorist functioning. These may be called methods based on terrorist observation and general user obstruction, respectively.

3.3.1 Apply Methods Based on Terrorist Observation

In any given terrorist activity, the terrorists involved use specific tactics and methods. Learning these specific tactics and methods may suggest ways to defend against them. For example, some researchers, through studying actual hackers with *qualitative research techniques*, have successfully identified specific tactics and methods that are relevant to cyberterrorism [154], a research strategy that may suggest methods of defence. More generally, it may be possible to use such research techniques with terrorists in custody, to identify specific terrorist approaches against which defences may be devised.

3.3.2 Apply Principles of Operator Obstruction

In some ways, terrorists are a distorted mirror image of intelligence and security personnel. Defence personnel function in teams; terrorists function in cells. Defence personnel gather intelligence; so do terrorists. Defence personnel engage in complex planning and decision-making; terrorists do the same. In short, terrorists face some of the same sort of challenges for effective cognitive and behavioral functioning that defence personnel do. In turn, this means that the same sorts of psychological research that is relevant to enhancing the functioning of defence personnel is relevant to impeding the functioning of terrorists. It is a matter of applying the same research—but in reverse. We may refer to this as applying principles of *operator obstruction*.

There are certain situations in which scientists may wish to prevent or impede personal functioning, rather than enhance that functioning [155]. For example, we accept the need to devise drug containers that are difficult to open—for children. Likewise, we exert much effort to make computer networks less accessible to hackers, at the same time that we try to make them more accessible to authorized users. Similarly, the many principles that HF/E scientists use to enhance performance can be inverted, as it were, to make terrorists and their cells less effective. Although specific examples are outside the scope of this document and its security classification, defence officials would be well advised to consult with specialists in *HF/E* and *I/O* psychology for help in devising programs to obstruct terrorist functioning.

4.0 PSYCHOLOGY AND CONSEQUENCE MANAGEMENT

At least into the foreseeable future, it seems likely that, despite our best efforts at prevention, some terrorist acts will occur. Consequence management involves measures to (1) limit the consequences of terrorist attacks, and (2) stabilize the situation in the aftermath of such attacks, in support of civilian authorities. Psychological methods can be applied to consequence management to enhance capabilities of first responders to terrorist incidents, to improve escape and evacuation procedures, to promote general resilience, and to treat victims effectively.

4.1 Enhance Capabilities of First Responders to Terrorist Incidents

We have already noted the relevance of research into *situation awareness (SA)* and *situated cognition (SC)* to the matter of improving the performance of counter-terrorism personnel. These same bodies of research are relevant to first responders to terrorist incidents (i.e., police, fire, and emergency medical personnel). Much like civil aviation and military personnel, for whom the concepts of SA and SC were first developed, the effectiveness of first responders depends upon their accurate awareness and understand of critical aspects of their environment, with life and death consequences. Improving the situation awareness and situated cognition of first responders should improve their effectiveness in responding to terrorist incidents. In addition, research into *decision-making* [73-79] is applicable to first responders, as it is to counter-terrorism personnel. HF/E

scientists have amassed a large body of research relevant to improving the function of first responders.

First responders, of course, typically function as members of teams. We also noted earlier the relevance of research regarding *team function* to the matter of improving the performance of counter-terrorism personnel [110-120]. This same body of research is relevant to improving the performance of first responders.

Earlier we noted the need for counter-terrorism personnel to have sophisticated tools. The same need exists for first responders. HF/E researchers have participated in the creation of such software tools as the Enhanced Consequence Management Planning and Support System (ENCOMPASS), which supports the information needs of multiple distributed first responder agencies, and the Domestic Emergency Response Information System, which supports emergency responses among multiple organizations responding to a crisis [1].

4.2 Improve Escape and Evacuation Procedures for Civilians

Post-9/11, there has been a renewal of research interest regarding how people react during fires and other occasions requiring building evacuations and emergency egress [1, 156]. *HF/E* researchers have long been concerned with the design of effective emergency signage and alarms for a civilian population [157, 158]. This research is highly relevant to the matter of designing escape and evacuation procedures for those present at the site of a terrorist attack. (Incidentally, researchers in these fields have also studied how to make alarms for first responders and military personnel more effective [159, 160].)

4.3 Promote Resilience in the General Population

One way to manage the consequences of terrorism is to take a proactive position: take steps to make the general population less vulnerable to some of the psychological effects of terrorism. This involves applying research regarding *resilience*, a topic focused upon by the subdiscipline of *positive psychology*. Resilience is a normal human process [161], but the capacity for resilience is not equally distributed across the population; resilience may vary with personal worldviews [121]. In recent years, researchers in positive psychology have made a good start at conducting research concerning human strengths [162-165], resistance to depression [166], inner "toughness" in the face of adversity [167, 168], and resilience [169-171], including specifically resilience in the face of terrorism [172-173]. Stress exposure training has been recommended for the military [177]; it may well be useful for the general population as well. This research is applicable to the construction of large-scale programs to promote the development of resilience among the general population. The effect of such an effort would be to make the population less vulnerable to some of the negative psychological effects of terrorist attack.

4.4 Treat Victims Effectively

We have come to understand much concerning the negative effects of a terrorist attack on mental and emotional health [178-181]. The American Psychological Association has provided information to the public to help people to cope with the aftermath of terrorism [182, 183]. Victims of terrorist attacks should be treated with appropriate *post-traumatic and crisis psychological treatment*. Psychology now has theoretical frameworks for understanding the psychological effects of terrorism (e.g., [184]). Although we yet have much to learn about appropriate treatment following 9/11-style catastrophes, we have some clinical advice for community-level intervention [185]; some effective and promising treatments for post-traumatic stress disorder in general have been identified [186, 187; cp. 188], including *virtual reality therapy* [189]. Research has also addressed treatment of military personnel following terrorist attack [190]. This entire research and

clinical literature is applicable to the effective treatment of victims of terrorist attack.

Finally, some clinical advice is available to help practitioners cope with the challenges of working with victims of trauma other than terrorism [191]. It may be expected that many of these guidelines will apply to practitioners treating victims of terrorism.

5.0 CONCLUSION AND OUTLOOK

We have here indicated the multiple ways in which various psychological subdisciplines can and indeed should contribute to DAT. There are several ways in which readers can apply this knowledge.

On one level, it would be appropriate to implement psychological expertise within a specific program, department, or government ministry. For example, it would be appropriate to call in psychological consultants to consider the functioning of a specific counter-terrorism program. On a wider scale, it would be appropriate to form a core of psychologists to consult on national or regional DAT as a whole. Our sense of the situation is that, the wider the scale of psychological intervention, the greater the benefit.

A comprehensive team of psychologists addressing DAT should be composed of representatives of several subdisciplines. Ideally, these would include specialists in *human factors and ergonomics* (useful in many areas, as noted above), *industrial/organizational psychology* (particularly specialists in team assessment and enhancement), *psychometrics* (to develop appropriate assessment instruments), *educational psychology* (to develop and assess instructional programs), *social and personality psychology* (specifically specialists on influence and persuasion), *cognitive psychology* (for cognitive enhancement), *clinical psychology* (for therapeutic intervention), and *counseling psychology* (specifically specialists in positive psychology and resilience). A team meant to address the issues we bracketed earlier (that is, the psychosocial origins of terror) should include specialists in *social and personality psychology* who are focused on those issues. In either case, the team leader should be someone with qualifications in *general psychology*; this possibly misleading designation refers to a special approach in psychology that is focused on unifying psychology, building bridges between and among the many specialties and subdisciplines within psychology [192].

6.0 REFERENCES

[1] American Psychological Association, Combatting terrorism: Responses from the behavioral sciences, 2004, retrieved from http://www.apa.org/ppo/issues/svignetteterror2.html on August 6, 2004.

[2] M. Crenshaw, “Terrorism,” A.E. Kazdin (Ed.), Encyclopedia of psychology: Vol. 8, Washington, DC: American Psychological Association, 2000, pp. 44-47.

[3] R.A. Hudson, The sociology and psychology of terrorism: Who becomes a terrorist and why?, Washington, DC: Library of Congress, Federal Research Division, 1999.

[4] F.M. Moghaddam, A.J. Marsella (Eds.), Understanding terrorism: Psychosocial roots, consequences, and interventions, Washington, DC: American Psychological Association, 2004.

[5] S.L. Plous, P.G. Zimbardo, “How social science can reduce terrorism,” The Chronicle of Higher Education, vol. 51, no. 3, September 10, 2004, pp. B9-B10.

[6] T. Pyszczynski, S. Solomon, J. Greenberg, In the wake of 9/11: The psychology of terror, Washington, DC: American Psychological Association, 2003.

[7] W. Reich (Ed.), Origins of terrorism: Psychologies, ideologies, theologies, states of mind (rev. ed.), Washington, DC: Woodrow Wilson Center Press, 1998.

[8] A. Silke (Ed.), Terrorists, victims and society: Psychological perspectives on terrorism and its consequences, West Sussex, England, UK: Wiley, 2003.

[9] N. Smelser, F. Mitchell (Eds.), Discouraging terrorism: Some implications of 9/11, Washington, DC: National Academies Press, 2002.

[10] N.J. Smelser, F. Mitchell (Eds.), Terrorism: Perspectives from the behavioral and social sciences, Washington, DC: National Academies Press, 2002.

[11] C.E. Stout (Ed.), The psychology of terrorism (4 vols.), Westport, CT: Praeger, 2002.

[12] Committee to Review the Security Design Criteria of the Interagency Security Committee, ISC security design criteria for new federal office buildings and major modernization projects: A review and commentary, Washington, DC: National Academies Press, 2003.

[13] M.G. Helander, "The human factors profession," G. Salvendy (Ed.), Handbook of human factors and ergonomics (2nd ed.), New York: Wiley, 1997, pp. 3-16.

[14] J.E. Harrigan, "Architecture and interior design," G. Salvendy (Ed.), Handbook of human factors and ergonomics (2nd ed.), New York: Wiley, 1997, pp. 964-986.

[15] K.A. Franck, "Environmental design research," A.E. Kazdin (Ed.), Encyclopedia of psychology: Vol. 3, Washington, DC: American Psychological Association, 2000, pp. 213-217.

[16] P.A. Hancock, S.G. Hart, "Defeating terrorism: What can human factors/ergonomics offer?," Ergonomics in Design, vol. 10, no. 1, Winter 2002, pp. 6-16.

[17] S. Kraemer, P. Carayon, R. Duggan, "Red team performance for improved computer security," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 1605-1609.

[18] J.E. Ormrod, Human learning: Theories, principles, and educational applications, New York: Merrill/ Macmillan, 1990.

[19] W.S. Terry, Learning and memory: Basic principles, processes, and procedures (2nd ed.). Boston: Allyn & Bacon, 2003.

[20] A. Schatz, "To secure airport, Houston enlists volunteer riders," The Wall Street Journal, vol. 244, no. 54, September 16, 2004, pp. A1, A6.

[21] P.A. Hancock, "Quo vadis homeland security," paper presented at the 46th annual meeting of the Human Factors and Ergonomics Society, Baltimore, Maryland, USA, 2002.

[22] R.J. Cohen, M.E. Swerdlik, Psychological testing and assessment (5th ed.), Boston: McGraw-Hill, 2002.

[23] J.C. Nunnally, I.H. Bernstein, Psychometric theory (3rd ed.), New York: McGraw-Hill, 1994.

[24] J.D. Bransford, A.L. Brown, R.R. Cocking (Eds.), How people learn: Brain, mind, experience, and school (expanded ed.), Washington, DC: National Academy Press, 2000.

[25] D. Druckman, R.A. Bjork (Eds.), Learning, remembering, believing: Enhancing human performance, Washington, DC: National Academy Press, 1994.

[26] S. Chaiken, W. Wood, A.H. Eagly, "Principles of persuasion," E.T. Higgins, A.W. Kruglanski (Eds.), Social psychology: Handbook of basic principles, New York: Guilford, 1996, pp. 702-742.

[27] R.B. Cialdini, Influence: Science and practice (4th ed.), Boston: Allyn & Bacon, 2000.

[28] R.H. Gass, J.S. Seiter, Persuasion, social influence, and compliance gaining (2nd ed.), Boston: Allyn & Bacon, 2002.

[29] E.S. Knowles, J.A. Linn (Eds.), Resistance and persuasion, Mahwah, NJ: Erlbaum, 2004.

[30] D.J. O'Keefe, Persuasion: Theory and research (2nd ed.), Thousand Oaks, CA: Sage, 2002.

[31] J.B. Stiff, P.A. Mongeau, Persuasive communication (2nd ed.), New York: Guilford, 2003.

[32] A.R. Andreasen, Marketing social change: Changing behavior to promote health, social development, and the environment, San Francisco: Jossey-Bass, 1995.

[33] D. Bornstein, How to change the world: Social entrepreneurs and the power of new ideas, Oxford, UK: Oxford University Press, 2004.

[34] R. Earle, The art of cause marketing: How to use advertising to change personal behavior and public policy, New York: McGraw-Hill, 2000.

[35] P. Kotler, N. Roberto, N. Lee, Social marketing: Improving the quality of life, Thousand Oaks, CA: Sage, 2002.

[36] N.K. Weinreich, Hands-on social marketing: A step-by-step guide, Thousand Oaks, CA: Sage, 1999.

[37] J.L. Fitzpatrick, J.R. Sanders, B.R. Worthen, Program evaluation: Alternative approaches and practical guidelines (3rd ed.), Boston: Allyn & Bacon, 2003.

[38] P.H. Rossi, M.W. Lipsey, H.E. Freeman, Evaluation: A systematic approach (7th ed.), Thousand Oaks, CA: Sage, 2003.

[39] J.S. Wholey, H.P. Hatry, K.E. Newcomer (Eds.), Handbook of practical program evaluation (2nd ed.), San Francisco: Jossey-Bass, 2004.

[40] M. Hall, "Airport screeners missed weapons," USA Today, September 23, 2004, p. 1A.

[41] The 9/11 Commission Report: Final report of the National Commission on Terrorist Attacks upon the United States (authorized edition), New York: Norton, 2004.

[42] P.A. Hancock, J.L. Szalma, "Vigilance and the price of freedom," Gateway: Human Systems Information Acquisition Center, vol. 13, no. 4, 2003, pp. 5, 20.

[43] M. W. Scerbo, "Stress, workload, and boredom in vigilance: A problem and an answer," P.A. Hancock, P.A. Desmond (Eds.), Stress, workload, and fatigue, Mahwah, NJ: Erlbaum, 2001, pp. 267-278.

[44] C.D. Wickens, J.G. Hollands, Engineering psychology and human performance (3rd ed.), Upper Saddle River, NJ: Prentice Hall, 2000.

[45] D.E. Broadbent, "Task combination and selective intake of information," Acta Psychologica, vol. 50, 1982, pp. 253-290.

[46] G. Matthews, D.R. Davies, S.J. Westerman, R.B. Stammers. Human performance: Cognition, stress and individual differences, Philadelphia: Taylor & Francis, 2000.

[47] J.E. Driskell, E. Salas (Eds.), Stress and human performance, Mahwah, NJ: Erlbaum, 1996.

[48] P.A. Hancock, P.A. Desmond (Eds.), Stress, workload, and fatigue, Mahwah, NJ: Erlbaum, 2001.

[49] T.D. Wickens, Elementary signal detection theory, Oxford, UK: Oxford University Press, 2002.

[50] J.F. Morgan, S. Stafford, R. Nayeem, A. Pepe, "Mental rotation of objects in an airport security setting," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 624-626.

[51] P.A. Hancock, J.L. Szalma, "The future of neuroergonomics," Theoretical Issues in Ergonomics Science, vol. 4, 2003, pp. 238-249.

[52] R. Parasuraman, "Neuroergonomics: Research and practice," Theoretical Issues in Ergonomic Science, vol. 4, 2003, pp. 5-20.

[53] C. Huff, "The baggage screener's brain scan," Monitor on Psychology, vol. 35, no. 8, September 2004, pp. 34-36.

[54] G. Belenky, "Sleep, performance, and network centric warfare," paper presented at the annual midyear symposium of American Psychological Association Divisions 21 and 19 and the Human Factors and Ergonomics Society Potomac Chapter, Fort Belvoir, Virginia, March 2004.

[55] J.A. Caldwell, Jr., J.L. Caldwell, D.L. Brown, J.K. Smith, "The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots," Military Psychology, vol. 16, 2004, pp. 163-181.

[56] N.L. Miller, L.G. Shattuck, "Sleep patterns and the impact on performance: A study of men and women enrolled at the United States Military Academy," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and

Ergonomics Society, 2004, pp. 1243-1247.

[57] M.R. Endsley, B. Bolté, D.G. Jones, Designing for situation awareness: An approach to user-centered design, London: Taylor & Francis, 2003.

[58] M.R. Endsley, "Design and evaluation for situation awareness enhancement," Proceedings of the Human Factors Society 32nd annual meeting, Santa Monica, CA: Human Factors Society, 1988, pp. 97-101.

[59] M.R. Endsley, "Situation Awareness Global Assessment Technique (SAGAT)" [NOR DOC 88-26], paper presented at the National Aerospace and Electronics Conference (NAECON), Dayton, Ohio; Hawthorne, CA: Northrop Corp., 1988.

[60] M.R. Endsley, "A methodology for the objective measurement of pilot situation awareness," paper in NATO Advisory Group for Aerospace Research and Development [AGARD] Conference Proceedings No. 478, Situational awareness in aerospace operations, 1989, pp. I1-I9.

[61] M.R. Endsley, "Toward a theory of situation awareness in dynamic systems," Human Factors, vol. 37, 1995, pp. 32-64.

[62] M.R. Endsley, D.J. Garland (Eds.), Situation awareness and measurement, Mahwah, NJ: Erlbaum, 2000.

[63] W.L. Shebilske, B.P. Goettl, D.J. Garland, "Situation awareness, automaticity, and training," M.R. Endsley, D.J. Garland (Eds.), Situation awareness analysis and measurement, Mahwah, NJ: Erlbaum, 2000, pp. 303-323.

[64] C.D. Wickens, "Workload and situation awareness," P.A. Hancock, P.A. Desmond (Eds.), Stress, workload, and fatigue, Mahwah, NJ: Erlbaum, 2001, pp. 443-450.

[65] D.A. Vincenzi, M. Mouloua, P.A. Hancock (Eds.), Human performance, situation awareness and automation: Current research and trends (2 vols.), Mahwah, NJ: Erlbaum, 2004.

[66] Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004.

[67] J. Eid, B.H. Johnsen, W. Brun, J.C. Laberg, J.K. Nyhus, G. Larsson, "Situation awareness and transformational leadership in senior military leaders: An exploratory study," Military Psychology, vol. 16, 2004, pp. 203-209.

[68] S.E. Graham, M.D. Matthews (Eds.), Infantry situation awareness: Papers from the 1998 Infantry Situation Awareness Workshop, U.S. Army Research Institute for the Behavioral and Social Sciences, 1999.

[69] M.D. Matthews, L.D. Strater, M.R. Endsley, "Situation awareness requirements for infantry platoon leaders," Military Psychology, vol. 16, 2004, pp. 149-161.

[70] J. Trahan, J. Hunter, "Situation awareness in the U.S. Army: A comparison of collaborative tools," paper presented at the annual midyear symposium of American Psychological Association Divisions 21 and 19 and the Human Factors and Ergonomics Society Potomac Chapter, Fort Belvoir, VA, March 2004.

[71] N.L. Miller, L.G. Shattuck, "A process model of situated cognition in military command and control," paper presented at the Command and Control Research and Technology Symposium, San Diego, California, May, 2004.

[72] L.G. Shattuck, N.L. Miller, "A process tracing approach to the investigation of situated cognition," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 658-662.

[73] G. Klein, "Applied decision making," P.A. Hancock (Ed.), Human performance and ergonomics, San Diego, CA: Academic Press, 1999, pp. 87-107.

[74] M.R. Lehto, "Decision making," G. Salvendy (Ed.), Handbook of human factors and ergonomics (2nd ed.), New York: Wiley, 1997, pp. 1201-1248.

[75] J.A. Cannon-Bowers, E. Salas (Eds.), Making decisions under stress: Implications for individual and team training, Washington, DC: American Psychological Association, 1998.

[76] R. Flin, E. Salas, M. Strub, L. Martin (Eds.), Decision making under stress, Burlington, VT: Ashgate, 1997.

[77] G. Klein, "The effect of acute stressors on decision making," J.E. Driskell, E. Salas (Eds.), Stress and human performance, Mahwah, NJ: Erlbaum, 1996, pp. 49-88.

[78] E. Salas, G. Klein (Eds.), Linking expertise and naturalistic decision making, Mahwah, NJ: Erlbaum, 2001.

[79] C. E. Zsambok, G. Klein (Eds.), Naturalistic decision making, Mahwah, NJ: Erlbaum, 1997.

[80] W.C. Elm, M.J. Cook, F.L. Greitzer, R.R. Hoffman, B. Moon, S.G. Hutchins, "Designing support for intelligence analysts," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 406-410.

[81] M.J. Cook, C.S. Angus, C.S.G. Adams, "Human factors of intelligence analysis," D.A. Vincenzi, M. Mouloua, P.A. Hancock (Eds.), Human performance, situation awareness and automation: Current research and trends: vol. 2, 2004, pp. 286-293.

[82] D.A. Thurman, W.D. Gray, "Cognitive factors in homeland defense: The role of human factors in the Novel Intelligence from Massive Data (NIMD) program," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 355-356.

[83] W.D. Gray, M.J. Schoelles, V.D. Veksler, "SimBorgs and simulated task environments for engineering next generation workstations for intelligence analysts," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 362-366.

[84] E. Whitaker, R. Simpson, L. Burkhart, R. MacTavish, C. Lobb, "Reusing intelligence analysts' search

plans," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 367-370.

[85] H. Nguyen, E. Santos, Jr., Q. Zhao, H. Wang, "Capturing user intent for information retrieval," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 371-375.

[86] A.R. Chappell, A.J. Cowell, D.A. Thurman, J.R. Thomson, "Supporting mutual understanding in a visual dialogue between analyst and computer," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 376-380.

[87] D.A. Boehm-Davis, R.W. Holt, R. Chong, J.T. Hansberger, "Using cognitive modeling to understand crew behavior," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 99-103.

[88] E.S. Connors, P.L. Craven, M.D. McNeese, T. Jefferson, Jr., P. Bains, D.L. Hall, "An application of the AKADAM approach to intelligence analyst work," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 627-630.

[89] P.A. Hancock, D.D. Diaz, "Ergonomics as a foundation for a science of purpose," Theoretical Issues in Ergonomics Science, vol. 3, 2002, pp. 115-123.

[90] K. Smith, P.A. Hancock, "Situation awareness is adaptive, externally directed consciousness," Human Factors, vol. 37, 1995, pp. 137-148.

[91] A. Allport, "Visual attention," M. I Posner (Ed.), Foundations of cognitive science, Cambridge, MA: The MIT Press, 1993, pp. 631-682.

[92] J. Duncan, "Attention," R.A. Wilson, F.C. Keil (Eds.), The MIT encyclopedia of the cognitive sciences, Cambridge, MA: The MIT Press, 1999, pp. 39-41.

[93] M.W. Eysenck, M.T. Keane, "Attention and performance limitations," D.J. Levitan (Ed.), Foundations of cognitive psychology: Core readings, Cambridge, MA: The MIT Press, 2002, pp. 363-398.

[94] A.H.C. Van der Heijden, "Attention," W. Bechtel, G. Graham (Eds.), A companion to cognitive science, Malden, MA: Blackwell, 1998, pp. 121-128.

[95] J. Baron, Thinking and deciding (3rd ed.), Cambridge, UK: Cambridge University Press, 2000.

[96] E. Shafir. "Decision making," R.A. Wilson, F.C. Keil (Eds.), The MIT encyclopedia of the cognitive sciences, Cambridge, MA: The MIT Press, 1999, pp. 220-223.

[97] E. Shafir, A. Tversky, "Decision making," D.J. Levitan (Ed.), Foundations of cognitive psychology: Core readings, Cambridge, MA: The MIT Press, 2002, pp. 601-620.

[98] J.F. Yates, P.A. Estin, "Decision making," W. Bechtel, G. Graham (Eds.), A companion to cognitive

science, Malden, MA: Blackwell, 1998, pp. 186-196.

[99] P. Cheng, "Causal reasoning," R.A. Wilson, F.C. Keil (Eds.), The MIT encyclopedia of the cognitive sciences, Cambridge, MA: The MIT Press, 1999, pp. 106-108.

[100]E. Shafir, "Probabilistic reasoning," R.A. Wilson, F.C. Keil (Eds.), The MIT encyclopedia of the cognitive sciences, Cambridge, MA: The MIT Press, 1999, pp. 671-672.

[101]J.P. Leighton, R.J. Sternberg (Eds.), The nature of reasoning, Cambridge, UK: Cambridge University Press, 2004.

[102]L.J. Rips, "Reasoning," W. Bechtel, G. Graham (Eds.), A companion to cognitive science, Malden, MA: Blackwell, 1998, pp. 299-305.

[103]L.J. Rips, "Deductive reasoning," R.A. Wilson, F.C. Keil (Eds.), The MIT encyclopedia of the cognitive sciences, Cambridge, MA: The MIT Press, 1999, pp. 225-226.

[104]P. Thagard, "Induction," R.A. Wilson, F.C. Keil (Eds.), The MIT encyclopedia of the cognitive sciences, Cambridge, MA: The MIT Press, 1999, pp. 399-400.

[105]P.A. Hancock, A.J. Masalonis, R. Parasuraman, "On the theory of fuzzy signal detection: Theoretical and practical considerations," Theoretical Issues in Ergonomic Science, vol. 1, 2000, pp. 207-230.

[106]R. Parasuraman, A.J. Masalonis, P.A. Hancock, "Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance," Human Factors, vol. 42, 2000, pp. 636-659.

[107]M.I. Jordan, S. Russell, "Computational intelligence," R.A. Wilson, F.C. Keil (Eds.), The MIT encyclopedia of the cognitive sciences, Cambridge, MA: The MIT Press, 1999, pp. lxxiii-xc.

[108]A.B. Badiru, J.Y. Cheung. Fuzzy engineering expert systems with neural network applications, New York: Wiley, 2002.

[109]M.E. Koltko-Rivera, "Detection of terrorist preparations by an artificial intelligence system employing fuzzy signal detection," Systems, concepts and integration (SCI) methods and technologies for defence against terrorism: Proceedings of the NATO Research and Technology Agency SCI-158 Symposium, 25-27 October 2004.

[110]M.T. Brannick, E. Salas, C. Prince (Eds.), Team performance and measurement: Theory, methods, and applications, Mahwah, NJ: Erlbaum, 1997.

[111]C.S. Burke, K.A. Wilson, E. Salas, "Teamwork at 35,000 feet: Enhancing safety through team training," Human Factors and Aerospace Safety, vol. 3, 2003, pp. 287-312.

[112]D. Frey, E. Jonas, T. Greitemeyer, "Intervention as a major tool of a psychology of human strengths: Examples from organizational change and innovation," L.B. Aspinwall, U.M. Staudinger (Eds.), A psychology of human strengths, Washington, DC: American Psychological Association, 2003, pp. 149-164.

[113]H.A. Priest, S. Burke, E. Salas, C.A. Bowers, "What do we know about teams? A practical framework,"

poster presented at the annual meeting of the American Psychological Association, Honolulu, Hawaii, July 2004.

[114]H.A. Priest, J. Guthrie, C.S. Burke, C. Bowers, E. Salas, "Are there differences in teamwork across team type?," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 2582-2586.

[115]E. Salas, C.A. Bowers, E. Edens (Eds.), Improving teamwork in organizations: Applications of resource management training, Mahwah, NJ: Erlbaum, 2001.

[116]S.H. Godar, S.P. Berris (Eds.), Virtual and collaborative teams: Process, technologies and practice, Hershey, PA: Idea Group Publishing, 2004.

[117]T. Jefferson, Jr., L. Ferzandi, M. McNeese, "Impact of hidden profiles on distributed cognition in spatially distributed decision-making teams," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 649-652.

[118]H.A. Priest, J. Guthrie, K.C. Stagl, C.S. Burke, E. Salas, "Virtual teams: The role of involvement, immersion and presence for distributed team members," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 2608-2612.

[119]A. Rittman, B. Wiggins, "Individual differences and considerations for dispersed collaborative performance," paper presented at the annual midyear symposium of American Psychological Association Divisions 21 and 19 and the Human Factors and Ergonomics Society Potomac Chapter, Fort Belvoir, VA, March 2004.

[120]E. Salas, S.M. Fiore (Eds.), Team cognition: Understanding the factors that drive process and performance, Washington, DC: American Psychological Association, 2004.

[121]M.E. Koltko-Rivera, "The psychology of worldviews," Review of General Psychology, vol. 8, 2004, pp. 3-58.

[122]J.E. Driskell, E. Salas, "Collective behavior and team performance," Human Factors, vol. 34, 1992, pp. 277-288.

[123]J.E. Driskell, E. Salas, J. Johnston, "Does stress lead to a loss of team perspective?," Human Performance in Extreme Environments, vol. 5, 2000, pp. 69-76.

[124]M.E. Koltko-Rivera, H.C.N. Ganey, J. Dalton, P.A. Hancock, "Worldview and acculturation as predictors of performance: Addressing these variables in human factors/ergonomics research," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 1223-1227.

[125]S.L. Rynes, D.M. Cable, "Recruitment research in the twenty-first century," W.C. Borman, D.R. Ilgen, R.J. Klimoski (Eds.), Handbook of psychology: Volume 12: Industrial and organizational psychology,

Hoboken, NJ: Wiley, 2003, pp. 55-76.

[126]N. Schmitt, J.M. Cortina, M.J. Ingerick, D. Wiechmann, "Personnel selection and employee performance," W.C. Borman, D.R. Ilgen, R.J. Klimoski (Eds.), Handbook of psychology: Volume 12: Industrial and organizational psychology, Hoboken, NJ: Wiley, 2003, pp. 77-105.

[127]J. Hogan, M. Lesser, "Selection of personnel for hazardous performance," J.E. Driskell, E. Salas (Eds.), Stress and human performance, Mahwah, NJ: Erlbaum, 1996, pp. 195-222.

[128]M.R. Barrick, A.M. Ryan (Eds.), Personality and work: Reconsidering the role of personality in organizations, San Francisco: Jossey-Bass, 2003.

[129]B.W. Roberts, R. Hogan (Eds.), Personality psychology in the workplace, Washington, DC: American Psychological Association, 2001.

[130]J.P. Campbell, D.J. Knapp (Eds.), Exploring the limits in personnel selection and classification, Mahwah, NJ: Erlbaum, 2001.

[131]S.K. Campbell, V.M. Catano, "Using measures of specific abilities to predict training performance in Canadian Forces operator occupations," Military Psychology, vol. 16, 2004, pp. 183-201.

[132]M.G. Rumsey, C.B. Walker, J.H. Harris (Eds.), Personnel selection and classification, Mahwah, NJ: Erlbaum, 1994.

[133]B. Schaab, J.D. Dressel, "What characteristics do digitally competent [i.e., cybercompetent] soldiers have in common?," D.A. Vincenzi, M. Mouloua, P.A. Hancock (Eds.), Human performance, situation awareness and automation: Current research and trends (2 vols.), Mahwah, NJ: Erlbaum, 2004, pp. 259-260.

[134]K. Kraiger, "Perspectives on training and development," W.C. Borman, D.R. Ilgen, R.J. Klimoski (Eds.), Handbook of psychology: Volume 12: Industrial and organizational psychology, Hoboken, NJ: Wiley, 2003, pp. 171-192.

[135]H.M. Cuevas, S.M. Fiore, C.A. Bowers, E. Salas, "Using guided learner-generated instructional strategies to transform learning into a constructive cognitive and metacognitive activity," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 1049-1053.

[136]E. Salas, C.A. Bowers (Eds.), "Training evaluation" (special issue), Military Psychology, vol. 15, no. 1, 2003.

[137]A. Bouzouane, L.-M. Gagnier, C. Dionne, I. Stiharu-Alexe, D. Gagne, "Toward virtual role playing based on intelligent agents," Applied Artificial Intelligence, vol. 13, 1999, pp. 777-794.

[138]S.G. Schiflett, L.R. Elliott, E. Salas, M.D. Coovert (Eds.), Scaled worlds: Development, validation and applications, Aldershot, Hampshire, UK: Ashgate Publishing, 2004.

[139]R.E.T. Jones, M.D. McNeese, E.S. Connors, T. Jefferson, Jr., D.L. Hall, "A distributed cognition simulation involving homeland security and defense: The development of NeoCITIES," Proceedings of

the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 631-634.

[140]K.M. Stanney (Ed.), Handbook of virtual environments: Design, implementation, and applications, Mahwah, NJ: Erlbaum, 2002.

[141]E. Salas, R.L. Oser, J.A. Cannon-Bowers, E. Daskarolis-Kring, "Team training in virtual environments: An event-based approach," K.M. Stanney (Ed.), Handbook of virtual environments: Design, implementation, and applications, Mahway, NJ: Erlbaum, 2002, pp. 873-892.

[142]B.W. Knerr, R. Breaux, S.L. Goldberg, R.A. Thurman, "National defense," K.M. Stanney (Ed.), Handbook of virtual environments: Design, implementation, and applications, Mahway, NJ: Erlbaum, 2002, pp. 857-872.

[143]T.A. Grotzer, D.N. Perkins, "Teaching intelligence: A performance conception," R.J. Sternberg (Ed.), Handbook of intelligence, Cambridge, UK: Cambridge University Press, 2000, pp. 492-515.

[144]R.J. Sternberg, Successful intelligence, New York: Simon & Schuster, 1996.

[145]R.J. Sternberg, Wisdom, intelligence, and creativity synthesized, Cambridge, UK: Cambridge University Press, 2003.

[146]V.J. Cassandro, D.K. Simonton, "Creativity and genius," C.L.M. Keyes, J. Haidt (Eds.), Flourishing: Positive psychology and the life well-lived, Washington, DC: American Psychological Association, 2003, pp. 163-183.

[147]J. Nakamura, M. Csikszentmihalyi, "The motivational sources of creativity as viewed from the paradigm of positive psychology," L.B. Aspinwall, U.M. Staudinger (Eds.), A psychology of human strengths, Washington, DC: American Psychological Association, 2003, pp. 257-269.

[148]R.S. Nickerson, "Enhancing creativity," R.J. Sternberg (Ed.), Handbook of creativity, Cambridge, UK: Cambridge University Press, 1999, pp. 392-430.

[149]D.K. Simonton, "Creativity," C.R. Snyder, S.J. Lopez (Eds.), Handbook of positive psychology, Oxford, UK: Oxford University Press, 2002, pp. 189-201.

[150]R.M. Hogarth, Educating intuition, Chicago: University of Chicago Press, 2001.

[151]P.B. Baltes, A.M. Freund, "Human strengths as the orchestration of wisdom and selective optimization with compensation," L.B. Aspinwall, U.M. Staudinger (Eds.), A psychology of human strengths, Washington, DC: American Psychological Association, 2003, pp. 23-35.

[152]P.B. Baltes, J. Glück, U. Kunzmann, "Wisdom: Its structure and function in regulating successful life span development," C.R. Snyder, S.J. Lopez (Eds.), Handbook of positive psychology, Oxford, UK: Oxford University Press, 2002, pp. 327-347.

[153]C. Carter, V. Gilligan, J. Shiban, F. Spotnitz (Writers), R. Bowman (Director), "Pilot" [television series episode], in C. Carter (Producer), The lone gunmen [Series], Los Angeles, CA: FOX/Ten Thirteen

Productions, episode first broadcast in the United States on 4 March 2001.

[154]T. Stanard, W.R. Lewis, D.A. Cox, D.A. Malek, J. Klein, R. Matz, "An exploratory qualitative study of computer network attacker cognition," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 401-405.

[155]P.A. Hancock, "Should human factors prevent or impede access?," Ergonomics in Design, vol. 6, no. 1, January 1998, p. 4.

[156]L. Winerman, "Fighting fire with psychology," Monitor on Psychology, vol. 35, no. 8, September 2004, pp. 28-30.

[157]K.R. Laughery, Sr., M.S. Wogalter, "Warnings and risk perception," G. Salvendy (Ed.), Handbook of human factors and ergonomics (2nd ed.), New York: Wiley, 1997, pp. 1174-1197.

[158]C. Mayhorn, M.S. Wogalter, J.L. Bell, "Are we ready? Misunderstanding homeland security safety symbols," paper presented at the annual midyear symposium of American Psychological Association Divisions 21 and 19 and the Human Factors and Ergonomics Society Potomac Chapter, Fort Belvoir, VA, March 2004.

[159]E.A. Bustamante, B.L. Anderson, J.P. Bliss, "Effects of varying the threshold of alarm systems and task complexity on human performance and perceived workload," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 1948-1952.

[160]E. Bustamante, C.K. Fallon, J.P. Bliss, "Adapting signal detection theory for analyzing human response to alarm signals," paper presented at the annual midyear symposium of American Psychological Association Divisions 21 and 19 and the Human Factors and Ergonomics Society Potomac Chapter, Fort Belvoir, VA, March 2004.

[161]A.S. Masten, "Ordinary magic: Resilience processes in development," American Psychologist, vol. 56, 2001, pp. 227-238.

[162]L.G. Aspinwall, U.M. Staudinger (Eds.), A psychology of human strengths, Washington, DC: American Psychological Association, 2003.

[163]J.E. Gilham (Ed.), The science of optimism and hope, Philadelphia: Templeton Foundation Press, 2000.

[164]C.L.M. Keyes, J. Haidt (Eds.), Flourishing: Positive psychology and the life well-lived, Washington, DC: American Psychological Association, 2003.

[165]C.R. Snyder, S.J. Lopez (Eds.), Handbook of positive psychology, Oxford, UK: Oxford University Press, 2002.

[166]L.Y. Abramson, L.B. Alloy, B.L. Hankin, C.M. Clements, L. Zhu, M.E. Hogan, W.G. Whitehouse, "Optimistic cognitive styles and invulnerability to depression," J. E. Gilham (Ed.), The science of optimism and hope, Philadelphia: Templeton Foundation Press, 2000, pp. 75-98.

[167]R.A. Dienstbier, L. M. Pytlik Zillig, "Toughness," C.R. Snyder, S.J. Lopez (Eds.), Handbook of positive psychology, Oxford, UK: Oxford University Press, 2002, pp. 515-527.

[168]C.D. Ryff, B. Singer, "Ironies of the human condition: Well-being and health on the way to mortality," L.G. Aspinwall, U.M. Staudinger (Eds.), A psychology of human strengths, Washington, DC: American Psychological Association, 2003, pp. 271-287.

[169]R.C. Drugan, "The neurochemistry of stress resilience and coping: A quest for nature's own antidote to illness," J. E. Gilham (Ed.), The science of optimism and hope, Philadelphia: Templeton Foundation Press, 2000, pp. 57-71.

[170]A.S. Masten, M.-G.J. Reed, "Resilience in development," C.R. Snyder, S.J. Lopez (Eds.), Handbook of positive psychology, Oxford, UK: Oxford University Press, 2002, pp. 74-88.

[171]S. Nolen-Hoeksema, "Growth and resilience among bereaved people," J. E. Gilham (Ed.), The science of optimism and hope, Philadelphia: Templeton Foundation Press, 2000, pp. 107-127.

[172]C.D. Ryff, B. Singer, "Flourishing under fire: Resilience as a prototype of challenged thriving," C.L.M. Keyes, J. Haidt (Eds.), Flourishing: Positive psychology and the life well-lived, Washington, DC: American Psychological Association, 2003, pp. 15-36.

[173]American Psychological Association, Fostering resilience in response to terrorism, 2004, retrieved from http://www.apa.org/psychologists/resilience.html on August 6, 2004.

[174]E.H. Grotberg, "From terror to triumph: The path to resilience," C.E. Stout (Ed.), The psychology of terrorism: Volume 1: A public understanding, Westport, CT: Praeger, 2002, pp. 185-207.

[175]J.H. Johnston, J.A. Cannon-Bowers, "Training for stress exposure," J.E. Driskell, E. Salas (Eds.), Stress and human performance, Mahwah, NJ: Erlbaum, 1996, pp. 223-256.

[176]J.M. Orasanu, P. Backer, "Stress and military performance," J.E. Driskell, E. Salas (Eds.), Stress and human performance, Mahwah, NJ: Erlbaum, 1996, pp. 89-125.

[177]J.M. Ross, H.C. Ganey, R.S. Broadway, "Efficacy of stress exposure training on target acquisition in combat simulations," Proceedings of the Human Factors and Ergonomics Society 48th annual meeting (New Orleans, Louisiana), Santa Monica, CA: Human Factors and Ergonomics Society, 2004, pp. 2187-2191.

[178]Y. Danieli, B. Engdahl, W.E. Schlenger, "The psychosocial aftermath of terrorism," F.M. Moghaddam, A.J. Marsella (Eds.), Understanding terrorism: Psychosocial roots, consequences, and interventions, Washington, DC: American Psychological Association, 2004, pp. 223-246.

[179]G. Sprang, "The psychological impact of isolated acts of terrorism," A. Silke (Ed.), Terrorists, victims and society, West Sussex, England, UK: Wiley, 2003, pp. 133-159.

[180]C.E. Stout (Ed.), The psychology of terrorism: Volume 2: Clinical aspects and responses, Westport, CT: Praeger, 2002.

[181]M.G. Wessels, "Terrorism and the mental health and well-being of refugees and displaced people," F.M.

Moghaddam, A.J. Marsella (Eds.), Understanding terrorism: Psychosocial roots, consequences, and interventions, Washington, DC: American Psychological Association, 2004, pp. 247-263.

[182]American Psychological Association, Psychology in daily life: Coping with terrorism, 2001, retrieved from http://helping.apa.org/daily/terrorism.html on August 6, 2004.

[183]American Psychological Association, Resources on coping with traumatic events, 2004, retrieved from http://www.apa.org/practice/ptresources.html on August 6, 2004.

[184]J. Kauffman (Ed.), Loss of the assumptive world: A theory of traumatic loss, New York: Brunner-Routledge, 2002.

[185]R. Gist, B. Lubin (Eds.), Response to disaster: Psychosocial, community, and ecological approaches, Philadelphia: Brunner/Mazel, 1999.

[186]G.S. Everly, Jr., J.M. Lating, Personality-guided therapy for posttraumatic stress disorder, Washington, DC: American Psychological Association, 2004.

[187]E.B. Foa, T.M. Keane, M.J. Friedman (Eds.), Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies, New York: Guilford, 2000.

[188]J.M. Lohr, R.W. Montgomery, S.O. Lilienfeld, D.F. Tolin, "Pseudoscience and the commercial promotion of trauma treatments," R. Gist, B. Lubin (Eds.), Response to disaster: Psychosocial, community, and ecological approaches, Philadelphia: Brunner/Mazel, 1999, pp. 291-326.

[189]B.O. Rothbaum, L.F. Hodges, D. Ready, K. Graap, R.D. Alarcon, "Virtual reality exposure therapy for Vietnam veterans with posttraumatic stress disorder," Journal of Clinical Psychiatry, vol. 62, 2001, pp. 617-622.

[190]D. Witten, K. McCormack-Chen, "Using Traumatic Incident Reduction (TIR) to ameliorate PTSD in military personnel present during the September 11th attack on the Pentagon: Case histories, theory, and practice," paper presented at the annual midyear symposium of American Psychological Association Divisions 21 and 19 and the Human Factors and Ergonomics Society Potomac Chapter, Fort Belvoir, VA, March 2004.

[191]M. Walker, "Supervising practitioners working with survivors of childhood abuse: Counter transference; secondary traumatization and terror," Psychodynamic Practice, vol. 10, 2004, pp. 173-193.

[192]R.J. Sternberg (Ed.), Unity in psychology: Possibility or pipedream?, Washington, DC: American Psychological Association, 2004.

7.0 ACKNOWLEDGEMENTS

This work was supported in part by a grant from the U.S. Army's Army Research Organization (Grant No. DAAD 19-01-1-0621: Multiple University Research Initiative—Operator Performance Under Stress [MURI-OPUS]), Dr. P.A. Hancock, Principal Investigator. The views expressed in this article are those of the authors and do not necessarily reflect the official position of the U.S. Department of the Army, Department of Defense, Department of Homeland Security, or the U.S. Government.

MENTAL ROTATION OF OBJECTS IN AN AIRPORT SECURITY SETTING

Justin Fox Morgan, Shawn Stafford, Razia Nayeem, Aaron Pepe
University of Central Florida, Orlando, Florida

The role of mental rotation in the airport security screening task was examined. 13 participants were asked to detect a handgun which was shown embedded in an x-ray image similar to what a baggage screener would see. The weapon was shown to the participant in 8 different rotations for 2 different perspectives (facing right or left for each rotation). The results of the experiment produced the classic inverted V response time function curve. The data suggest that mental rotation can be applied to complex stimuli found in detection task such as the airport security domain.

INTRODUCTION

Much research has been done examining the role and applications of mental rotation since Shepard and Metzler's 1971 findings suggesting a process of mental rotation. Studies have allowed researchers to study how individuals detect objects that are shown to them at different rotations. This research, when applied, could play a crucial roll in helping observers such as airport baggage screeners.

Since the events of September 11th, 2001, it has become evident that airport security personal need a better, more efficient, system of training at detecting objects that pass through airport checkpoints. There has been little research in applying mental rotation to the detection of threatening objects. The current study looked at a simple handgun rotation task, and how observers detected the handgun shown to them at different rotations.

Mental rotation has been applied to basic stimuli such as tetris shapes (e.g., Sims & Mayer, 2002), stick figures (e.g., Uecker & Obrzut, 1993), simple shapes (e.g., Cohen & Dale, 1993), and line shapes (e.g., Tarr & Pinker, 1989), which typically produces the classic inverted V-shaped mental rotation response time function (Wexler, Kosslyn & Berthoz, 1998). It is hypothesized that the same response time function that is consistently found for basic stimuli would transfer to complex stimuli (a handgun) in the applied domain of airport baggage screening.

METHOD

Participants were 13 university college students between the ages of 19 and 30. Superlab for Windows was used to present the images to the participant and record the data. The experiment consisted of showing a Berretta 92 handgun in 8 different rotations (0º, 45º, 90º, 135º, 180º, 225º, 270º, 315º). The participants where shown the hand gun in two different orientations. The handgun faced right for half of the trials, and left for half of the trials, for a total of 240 images. Each presented image was followed by a 1000ms visual mask of the computer display.

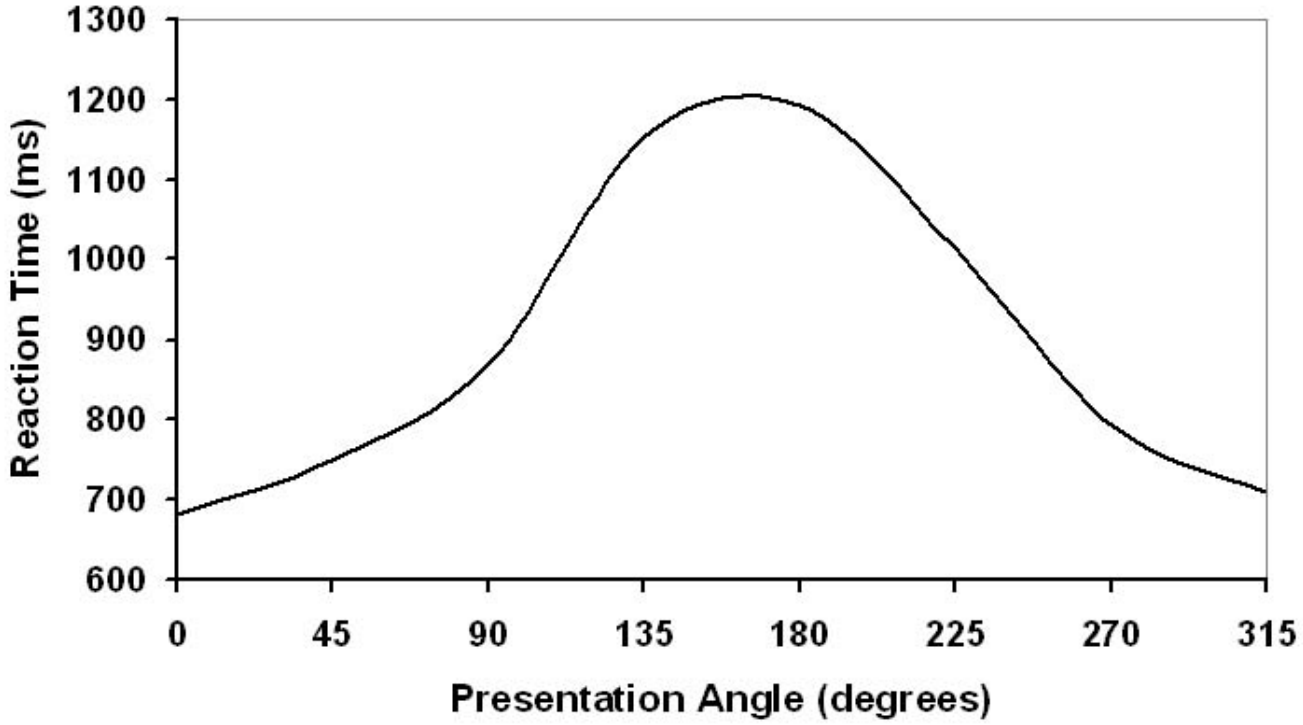
Participants viewed two handguns on each image. The handgun on the left of the screen was shown at a 0º rotation, facing to the right, for the duration of the experiment. The participants had to match one of 16 different handgun images on the right of the screen (8 rotations at 2 orientations), to the handgun on the left of the screen. If the handgun on the right side of the image is the same orientation as the handgun on the left, the participant had to press the zero key on a standard keyboard. If the two handguns were of different orientations the participant was told to press the enter key. Additionally, the handgun on the right side of the image was embedded in an airport baggage screening x-ray image. The x-ray image was of a suitcase with random objects (such as clothing, belts, and coat hangers) in the background. All images used the same suitcase x-ray, with the different oriented-rotated handguns inserted into the same position.

RESULTS

All data from each participant was screened to insure each participant exceeded a predefined level of accuracy (set at 80%). No participant data had to be excluded from the analysis. Reaction times (RT) were collected for each participant and mean reaction times were calculated for each rotation. Mean RT was collapsed across rotation for all participants. An ANOVA was conducted on the means of the 8 rotations. The main effect for reaction time was significant, $F(7, 77) = 21.28$, $p<.001$. For a participant's data to be included they had to score correctly hit rate of 80% minimum. All 13 participants scored above 80%.

Only a small subset of the data, RT's for the target matching orientation, is presented in this paper. A Bonferroni corrected paired-sample t-test was run on the data. The 0º rotation was used as the control and compared to all other rotations. The data showed the 0º right orientation was significantly different from the 90º, 135º, 180º, 225º, and 270º right orientation rotations at p<.007. Only at the 45º and 315º right orientation rotations was there no significant effect. The mean RT's for rotation when plotted (Figure 1) produce the classic inverted V-shaped response time function.

Figure 1: Reaction time as a function of presentation angle



DISCUSSION

The results support the hypothesis that mental rotation can be applied to the domain of airport security for the detection of some types of dangerous items. The results of this study predict that some rotations of weapons that pass through airport security checkpoints may be more difficult to detect by observers. Though this experiment only uses picture plane rotations of the handgun, it does implicate the need for further research into rotation along the z axis. However, it is obvious from the results of this experiment that further investigation is needed, improving the safety countless travelers.

REFERENCES

Cohen, Dale., Kubovy., 1993. Mental Rotation, Mental Representation, and Flat Slopes. *Cognitive Psychology*, 25, 351-382.

Sims, Valerie K., Mayer, Richard E., 2002. Domain Specificity of Spatial Expertise: The Case of Video Game Players. *Applied Cognitive Psychology, 16*, 97-115.

Shepard, R.N., Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, (171), 701-703.

Tarr, Michael J., Pinker, Steven., 1989. Mental Rotation and Orientation-Dependence in Shape Recognition. Cognitive Psychology, 21, 233-282.

Uecker, Anne., Obrzut, John E., 1993. Hemisphere and Gender Differences in Mental Rotation. *Brain and Cognition*, 22, 42-50.

Wexler, Mark., Kosslyn, Stephen M., Berthoz, Alain., 1988. Motor processes in mental rotation. *Cognition*, 68, 77-94.

Personality Theory and Human Factors Research

Mark E. Koltko-Rivera
Professional Services Group, Inc.

ABSTRACT:

This paper reflects on issues raised in Schaab's (2004) presentation concerning personality characteristics of the cyber-competent. Schaab's findings raise the possibility that personality traits affect cyber-competence, an insight that is certainly congruent with everyday experience, where personality is seen as affecting human performance in many ways. To apply personality theory to human factors domains, researchers have available to them a variety of theoretical frameworks to study traits (including factorial and circumplex models) and motives (including specific motive and motivational structure theories), for all of which operationalizations are available. There is also a pressing need to develop a set of scales to assess attitudes towards high technology. Human factors researchers should use these theoretical frameworks and operationalizations to study how personality moderates human interaction with the products of high technology (e.g., computers, robots, software agents); this would be the first step in learning how to enhance the cyber-competence of all people.

Keywords: Personality, motivation, human factors, digital competence, cyber-competence, cyber-performance, attitudes toward technology

I have been asked to respond to issues raised by the paper presented by Dr. Brooke Schaab (2004). I concentrate on why and how human factors research should focus on issues addressed by personality theory.

Dr. Schaab administered the Self-Directed Search (SDS; Holland, Powell, & Fritzsche, 1997) to 127 U.S. Army soldiers who had been selected to be trained as Army military analysts; these soldiers were to be trained to work with the Army's most advanced digital systems. The SDS is based on Holland's (1997) model of vocational personalities and work environments; this model posits six vocational personality dimensions, corresponding to six work environment dimensions (Realistic, Investigative, Artistic, Social, Enterprising, and Conventional); the theory proposes that a person with a given personality configuration would perform best in a job with a congruent work environment configuration. In her research, Dr. Schaab found that the overwhelming majority (98%) of the Army analysts-in-training had personality configurations that loaded highly on one or both of the Investigative or Realistic dimensions.

These results are at least compatible with the notion that digital competence (i.e., competence in working within a highly computerized environment) is not equally distributed across personality types; rather, some personality types are simply more digitally competent than others. Such a finding, if replicated, would have profound consequences for human factors theory, research, and practice.

The "Why" of Applying Personality Theory to Human Factors Research

Given the potential consequences, I find it interesting that Dr. Schaab's research was the only report presented at the HPSAA II conference that placed its primary focus upon the influence of personality on a human factors variable. It would appear that human factors research is still guided predominantly by the position of Fitts, who suggested over 40 years ago that personality is of little importance to human factors scientists and practitioners (Fitts, 1963, p. 924)

However dominant this position is in human factors research and practice, it is wildly incongruent with our experience of everyday life in the real world, where we all know that personality affects performance. This is one

reason why we assign some kinds of work to some people and not to others. Of course training and experience play a great part in moderating performance, but personality is an important moderator as well.

In the spirit of recognizing this issue, I would suggest that we extend the question raised by the title of Schaab's (2004) paper. Limiting myself to the domain of digital competence (or cyber-competence, as I think it is better designated), I suggest that two fruitful questions for human factors researchers to consider are the following:

- What personality characteristics are typical of more and less cyber-competent people? (I.e., how does personality moderate cyber-competence and cyber-performance?)

- How can we compensate for the personality characteristics of less cyber-competent people?

These are not small issues. Within military contexts, the move to network-centric warfare (Galster & Bolia, 2004) suggests that cyber-competence will be important to attain military objectives. Within civilian contexts, all indications suggest that cyber-competence is becoming increasingly important in successfully negotiating both the demands of everyday life and the demands of many work environments. Consequently, an understanding of how personality moderates cyber-competence and cyber-performance is important for enhancing human performance in many contexts. So, how might such research be pursued?

The "How" of Applying Personality Theory to Human Factors Research

Kurt Lewin noted that there is nothing so practical as a good theory. Human factors scientists have several choices when it comes to applying personality theory to the human factors research milieu. Personality theories and variables may be considered as falling into four classes: traits, motives, cognitions, and social context (Winter & Barenbaum, 1999). I will focus here on traits and motives. (My colleagues and I have dealt elsewhere with the issue how the effect of cognitions and social context on human factors variables can be approached, when we describe how theories of worldview and acculturation may be applied to human factors research; Koltko-Rivera, Ganey, Hancock, & Dalton, 2004. Concerning worldview, see also Koltko-Rivera, 2004.)

Trait Approaches to Personality

Trait theories construe personality as a collection or profile of dimensions or traits. These traits are often conceived in bipolar terms (e.g., optimism vs. pessimism). Two major classes of models of traits are *factorial models* and *circumplex models.*

Factorial models consider personality traits to be collected into larger factors. Probably the most popular factorial model currently is the Five Factor model of personality (McCrae & Costa, 1999), which collects dozens of personality traits into five overarching supertraits, which can be recalled by the acronym OCEAN: *O*penness to experience (vs. closedness to new things), *C*onscientiousness (vs. tendency to disorder), *E*xtraversion (vs. introversion), *A*greeableness (versus disagreeableness), and *N*euroticism (vs. mental healthiness). The five-factor approach to personality traits has a long history in personality research, and the five-factor structure seems to be replicable across many cultural contexts (John & Srivastava, 1999). The Revised NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992; Piedmont, 1998) offers one operationalization of the five-factor theory of personality, and has been used in many research projects. In addition, many instruments are available to assess individual traits or small groups of traits (e.g., Zuckerman & Lubin, 1985). (Of course, there are many instruments to assess psychopathology, which may be considered a superfacet of the Neuroticism factor of personality. For sake of brevity, I will mention only one, which addresses multiple aspects of psychopathology: the Personality Assessment Inventory; Morey, 1991, 2003.)

Circumplex models consider personality traits to be distributed along one or more circular spectra, like a color wheel. On such a circular spectrum, or circumplex, some traits appear close together (e.g., "sarcastic" and

"rebellious") while others appear on opposite sides of the circumplex (e.g., "arrogant" and "deferential"). Many circumplex models are possible, depending on the type of traits being studied (e.g., interpersonal traits, psychopathological traits, etc.); a variety of instruments are available to operationalize these constructs (see multiple papers in Plutchik & Conte, 1997).

Motivational Approaches to Personality

Theories of motivation tend towards two types. One we may call the *specific motive* theories, while the other we may consider as *motivational structure* theories.

Specific motive theories focus on specific motives or lists of motives. For example, research has focused on the need for achievement (McClelland, Atkinson, Clark, & Lowell, 1976) and the need for power (McClelland & Burnham, 1976).

Motivational structure theories focus on personality structures that have motivational consequences. For example, Maslow posited a hierarchy of motivations that must be addressed in a specific order, ranging from safety and security through self-actualization and self-transcendence (Maslow, 1969, 1970). The famous developmental sequence derived from psychoanalytic theory is also a motivational theory (defining oral, anal, phallic, and genital needs; Freud, 1940/1969, Chap. 3). An analytical psychology model, Jung's (1921/1971) theory of psychological types, may be construed as a model of motivation: extraverts are motivated to seek stimulation from the external world, introverts from the internal world; sensing types are motivated to seek data for decisions from the sensory world, while intuitive types are motivated to seek data for decisions from the world of intuitions; thinking types then are motivated to make decisions on the basis of linear logic, feeling types on the basis of emotional logic. The Multitheory Personality Assessment Instrument (Koltko-Rivera & Torres, 2004) provides operationalizations for these three models, the Maslovian, Freudian, and Jungian.

Concluding Remarks

When these remarks were shared at the HPSAA II conference, Dr. Christina Frederick-Recascino noted the following:

- The relationship of personality trait and motivation to performance may not be direct, but rather may be mediated by attitudes.

- There is a distinct need to educate human factors professionals in how to apply personality theory to human factors research and practice.

In relation to the first point, it is nothing short of scandalous that, at this late date, we have not developed a general purpose scale regarding attitudes towards higher technology. Anecdotal evidence suggests that there is a great deal of variation in these attitudes; although many people (including, I suspect, most people who inhabit desks near human factors scientists) have a positive and accepting attitude to high technology, many other people regard high technology with suspicion and even fear. Doubtless these attitudes (which may have trait and motivational underpinnings) affect human-computer interaction, and human interaction with any of the products of high technology.

In relation to the second point, this article and others (e.g., Ganey, Koltko-Rivera, Murphy, Hancock, & Dalton, 2004; Koltko-Rivera, Ganey, Hancock, & Dalton, 2004; Koltko-Rivera, Hancock, Ganey, & Dalton, 2004) are an attempt to educate human factors professionals about the need to consider personality theory (as well as theory regarding affect, worldview, and acculturation) in research and practice. This is an area that will only serve to enrich human factors research and practice.

Acknowledgement

The views expressed in this work are those of the author and do not necessarily reflect official Army policy. This research was supported in part by the Office of the Secretary of Defense, Grant DAMD17-02-C-0016 (Principal Investigator: Mark E. Koltko-Rivera), and by the U.S. Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office, Grant DAAD19-01-1-0621 (Principal Investigator: Peter A. Hancock).

References

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Psychological Assessment Resources: Lutz, FL.

Fitts, P. M. (1963). Engineering psychology. In S. Koch (Ed.), *Psychology: A study of a science: Vol. 5* (pp. 908-933). New York: McGraw-Hill.

Freud, S. (1969). *An outline of psycho-analysis*. New York: Norton. (Original work published 1940)

Galster, S. M., & Bolia, R. S. (2004, March). *Decision quality and mission effectiveness in a simulated command and control environment*. Paper presented at the Second Conference on Human Performance, Situation Awareness and Automation (HPSAA II), Daytona Beach, Florida.

Ganey, H. C. N., Koltko-Rivera, M. E., Murphy, L. M., Hancock, P. A., & Dalton, J. (2004, March). *Affect as an aspect of workload: Why and how to address affect in automation and simulation studies.* Poster presented at the Second Conference on Human Performance, Situation Awareness and Automation (HPSAA II), Daytona Beach, Florida.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Lutz, FL: Psychological Assessment Resources, Inc.

Holland, J. L., Powell, A. B., & Fritzsche, B. A. (1997). *Self-Directed Search (SDS) professional user's guide: 1994 edition*. Lutz, FL: Psychological Assessment Resources, Inc.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York: Guilford.

Jung, C. G. (1971). *The collected works of C. G. Jung. Vol. 6: Psychological types* (R. F. C. Hull, Trans.). New York: Princeton University Press. (Original work published 1921)

Koltko-Rivera, M. E. (2004). The psychology of worldviews. *Review of General Psychology, 8*, 3-58.

Koltko-Rivera, M. E., Ganey, H. C. N., Hancock, P. A., & Dalton, J. (2004). *Worldview and acculturation as aspects of workload: Why and how to address these factors in automation and simulation studies.* Paper proposed for conference presentation.

Koltko-Rivera, M. E., Hancock, P. A., Ganey, H. C. N., & Dalton, J. (2004). *An invitation to the dark side: Human factors professionals encounter personality, motivation, affect, worldview, and acculturation psychology.* Manuscript in preparation.

Koltko-Rivera, M. E., & Torres, N. (2004). *The Multitheory Personality Assessment Instrument—Freudian, Jungian, and Maslovian Scales (MPAI-FJM)*. Manuscript in preparation.

Maslow, A. H. (1969). The farther reaches of human nature. *Journal of Transpersonal Psychology, 1*, 1-9.

Maslow, A. H. (1970). *Motivation and personality* (2nd ed.). New York: Harper & Row.

McClelland, D. C., Atkinson, J. W., Clark, R. W., & Lowell, E. L. (1976). *The achievement motive* (2nd ed.). New York: Irvington.

McClelland, D. C., & Burnham, D. (1976, March-April). Power is the great motivator. *Harvard Business Review*, pp. 100-111.

McCrae, R. R., & Costa, P. T., Jr. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139-153). New York: Guilford.

Morey, L. C. (1991). *The Personality Assessment Inventory: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Morey, L. C. (2003). *Essentials of PAI assessment*. n.p.: Wiley.

Piedmont, R. L. (1998). *The Revised NEO Personality Inventory: Clinical and research applications*. New York: Kluwer Academic/Plenum.

Plutchik, R., & Conte, H. R. (Eds.). (1997). *Circumplex models of personality and emotions*. Washington, DC: American Psychological Association.

Schaab, B. (2004, March). *What personality characteristics do digitally competent soldiers have in common?* Paper presented at the Second Conference on Human Performance, Situation Awareness and Automation (HPSAA II), Daytona Beach, Florida.

Winter, D. G., & Barenbaum, N. B. (1999). History of modern personality theory and research. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 3-27). New York: Guilford.

Zuckerman, M., & Lubin, B. (1985). *Manual for the MAACL-R: The Multiple Affect Adjective Check List—Revised (MAACL-R)*. San Diego, CA: Educational and Industrial Testing Service (EdITS).

WORLDVIEW AND ACCULTURATION AS PREDICTORS OF PERFORMANCE: ADDRESSING THESE VARIABLES IN HUMAN FACTORS / ERGONOMICS RESEARCH

Mark E. Koltko-Rivera[1], H. C. Neil Ganey[2], Joseph Dalton[2], Peter A. Hancock[2,3]

[1]Professional Services Group, Inc., Winter Park, Florida

[2]MURI-OPUS Laboratory and Department of Psychology
[3]Institute for Simulation and Training
University of Central Florida
Orlando, Florida

Researchers in human factors and ergonomics (HF/E) are encouraged to consider two classes of predictor variables that typically are not included in traditional performance research: worldview and acculturation. Worldviews are sets of assumptions about life and the physical and social worlds. Acculturation involves the degree of commitment made to a culture's set of values and practices. Worldview and acculturation are highly relevant to performance. Two worldview variables, Individualism—Collectivism and Locus of Control, have attracted research interest. At least four other dimensions within Koltko-Rivera's (2004) collated model of worldview are likely to be relevant to performance: mutability, time orientation, relation to authority, and interaction. As industrialized societies become more diverse, acculturation becomes more relevant to performance. Several areas are identified for future research, such as worldview/acculturation—task interactions, and team cognition. Assessment instrumentation is briefly described.

Within the field of human factors and ergonomics (HF/E), research has focused on many areas of human performance: human-computer interaction, aerospace and automobile control, handicapped functioning, etc. (see Hancock, 1999). Despite the variety of areas investigated, HF/E research on performance typically uses only three categories of predictors (including both independent and moderating variables): task characteristics (e.g., vigilance demand), environmental stress (e.g., heat), and individual differences (e.g., personality). (See: Helander, 1997; Hockey & Hamilton, 1983.) Individual differences variables have attracted less attention, and that only fairly recently, perhaps as a reflection of the strong roots of HF/E within engineering and experimental psychology (Fitts, 1963). However, even including individual differences variables, a surprisingly large amount of performance variance is unaccounted for in most HF/E research.

As part of the Multidisciplinary University Research Initiative (MURI) at the University of Central Florida, we are developing a meta-theory of performance that posits the action of several neglected classes of predictor variables. We believe that these variables may account for much of the unaccounted-for variance in performance research (Koltko-Rivera, Ganey, Dalton, & Hancock, 2004). Among these emerging predictors are two classes of individual differences variables: worldview and acculturation. Although few studies in the HF/E literature have addressed these variables, several studies in other subdisciplines demonstrate the potential relevance of these variables for HF/E research. This paper defines worldview and acculturation, and explains their relevance to performance research. Our objective is to encourage HF/E researchers to use these potentially powerful classes of predictor variables. The spirit behind our objective is to more closely achieve that "integration . . . of psychosocial, cultural, and technological factors with human-machine performance interface factors" that has inspired the general sociotechnical systems approach ("Technical Group," 1984, p. vii).

WORLDVIEW

A worldview is a cognitive structure consisting of foundational beliefs. Metaphorically, it may be thought of as a lens through which the individual views reality:

> A worldview is a way of describing the universe and life within it, both in terms of what is and what ought to be. A given worldview is a set of beliefs that includes limiting statements and assumptions regarding what exists (either in actuality, or in principle), what objects or experiences are good or bad, and what objectives, behaviors, and relationships are desirable or undesirable. A worldview defines what can be known or done in the world, and how it can be known or done. In addition to defining what goals can be sought in life, a worldview defines what goals should be pursued. Worldviews include assumptions that may be unproven, and even unprovable, but these assumptions are superordinate, in that they provide the

> epistemic and ontological foundations for other beliefs within a belief system. (Koltko-Rivera, 2004, p. 4)

Koltko-Rivera's (2004) collated model of worldview structure posits 42 dimensions that have been identified in the literatures of psychology, anthropology, and elsewhere. Two of these dimensions have been the focus of a great deal of research: Locus of Control (a subset of the Control Location dimension in the collated model), and Individualism—Collectivism (called Relation to Group in the collated model).

Locus of Control / Control Location

Locus of control (Rotter, 1966) refers to beliefs about the source of the reinforcements or rewards that follow behavior. The belief that rewards result from a person's efforts or attributes reflects *internal* locus of control; the belief that these rewards are under the control of fate, chance, or social forces reflects *external* locus of control. Locus of control has been identified as influencing job stress (Center for Social Epidemiology, n.d.), and a wide variety of behaviors (Lefcourt, 1982), including job performance (Judge & Bono, 2001). It is reasonable to think that this worldview dimension might be relevant to any number of training and performance situations. For example, one would expect internal locus of control individuals to put more effort into training and task attainment than external locus of control individuals.

Some HF/E research has focused (even if indirectly) on locus of control in relation to performance. For example, locus of control correlates with laboratory tests of cognition and memory (Arbuckle, Gold, Andres, Schwartzman, & Chaikelson, 1992; Erez & Judge, 2001). Locus of control also correlates with self-efficacy and performance goal orientation (Phillips & Gully, 1997). Certainly this construct deserves further attention from HF/E researchers interested in correlates of performance.

Individualism—Collectivism / Relation to Group

Individualism—Collectivism (Kagitçibasi, 1997) refers to beliefs about what is appropriate to do, when one's personal agenda is in conflict with the agenda of one's reference group (e.g., one's family or work team). An *individualism* orientation states that one's personal agenda takes priority in such situations; a *collectivism* orientation states that the reference group's agenda takes priority. Individualism-collectivism has strong effects on multiple aspects of behavior (Kagitçibasi, 1997; Wheeler, Reis, & Bond, 1989). This worldview dimension might be relevant to many teamwork situations. For example, one might expect a team largely comprised of individualism-oriented individuals to perform less well on a team task than a team comprised largely of collectivism-oriented individuals.

Some HF/E research has focused (even if indirectly) on individualism—collectivism (I-C) and its relationship to performance. For example, I-C was related to conformity on Asch-type line judgment tasks (Bond & Smith, 1996). Stress-induced individualistic focus has been found to degrade team performance (Driskell, Salas, & Johnston, 2000). This dimension of worldview, too, deserves further attention from HF/E researchers interested in correlates of performance.

Other Worldview Dimensions of Interest for HF/E

Several other dimensions of the collated model of worldview are relevant to HF/E research.

Mutability. This worldview dimension involves beliefs about whether people can change longstanding characteristics, such as habits or character. Someone with a *changeable* orientation believes that people can so change, while someone with a *permanent* orientation on this dimension believes that people cannot. Mutability beliefs are relevant to training situations. We might expect people with a changeable orientation to make a relatively better adjustment in training situations that have the objective of thoroughgoing transformation, such as basic training in the military; on the other hand, we might expect those with a permanent orientation to make a relatively poorer adjustment.

Time orientation. This dimension refers to beliefs about the appropriate focus of one's activities: to honor the ways of the *past*, to experience the possibilities of the *present*, or to focus on achievement in the *future*. Time orientation beliefs are relevant to many performance situations. In general, we would expect future time orientation to be associated with goal-achievement. However, this is an area in which respect for cultural differences is crucial; the future-oriented team leader may find it useful to position goal achievement in terms of upholding the honorable traditions of the past, for team members who are past-oriented.

Relation to authority. This dimension refers to preferences for different authority structures. Individuals at the *linear* pole prefer well-defined hierarchical differences, while individuals at the *lateral* pole prefer egalitarian structures. Beliefs about relation to authority are relevant to the function of many structured group situations, such as the military. We would expect work groups in which there is consistency of beliefs on relation to authority across team members and leaders to function more smoothly and effectively than groups where there are clashing beliefs on this dimension. It is possible that individuals and authority-worldview-homogeneous groups who function from a linear position will perform differently than those who function from a lateral position; it is also possible that for some types of tasks the linear position will result in superior performance, while for other tasks the lateral position will do so. (We say more below concerning the interaction of task type with homogeneity.)

Interaction. This dimension reflects one's preferred manner of interacting with others, whether that

preference be for *competition* with, *cooperation* with, or *disengagement* from others. This aspect of worldview is relevant to many aspects of group function. One set of hypotheses involves homogeneity of worldview: we might expect groups that are homogeneous in terms of the interaction dimension to perform differently than groups that are heterogeneous with respect to this aspect of worldview. Another set of hypotheses involves the differential effects of different worldview positions. For example, we might expect individuals with competitive, cooperative, or disengaging preferences to perform differently in group situations; in addition, we might expect groups that are heterogeneously competitive, cooperative, or disengaging to perform differently on group tasks. (The use of the term "perform differently" is deliberately vague; at the current state of knowledge, different theoretical positions on group function would lead to different directional hypotheses.)

ACCULTURATION

Acculturation reflects the extent to which someone has a commitment to a culture's mode of expression and valuation. Thus, if someone has a strong commitment to American middle-class values, we would say that this person is strongly acculturated to middle-class America – regardless of the person's citizenship or nationality. Acculturation can be studied on both group and individual levels (Berry & Sam, 1997). Acculturation is at least bidimensional (Ryder, Alden, & Paulhus, 2000), and has complex affective, behavioral, and cognitive aspects (Chun, Balls Organista, & Marín, 2003).

We have every reason to believe that acculturation has wide-ranging effects on performance. This is so because acculturation has significant effects on multiple domains of human experience, such as family functioning, mental health, and physical health (Chun et al., 2003). Acculturation correlates with marital conflict (Flores, Tschann, Marin, & Pantoja, 2004); we should not be surprised if it correlates with conflict in work teams. Acculturation correlates with different patterns of anger regulation and attachment in family relationships (Kim & Zane, 2004); similarly, we should not be surprised if it correlates with anger regulation and attachment in teams. Acculturation correlates with patterns of emotional response (Liem, Lim, & Liem, 2000); we may see the same in work groups, with concomitant effects on performance.

Acculturation correlates (through ethnic identity) with the ultimate inhibitor of performance: suicide (Kaslow et al., 2004). Given that an unexplained spike in suicides among troops in Iraq is a matter of serious recent concern to the U.S. Department of Defense (Zoroya, 2003), factors impinging on suicidality are a reasonable concern for HF/E researchers.

It is hard to imagine a world in which acculturation would *not* affect performance, given the degree to which values—including work values—are bound up with culture. There are systematic and wide differences between cultures' values (Smith & Schwartz, 1997). As societies become more ethnically and culturally diverse, the degree of an individual's acculturation to a given society's work values becomes highly relevant to studies of performance. Acculturation is a major precursor to social adaptation (Berry & Sam, 1997), including adaptation to the workplace. For example, acculturation plays a role in the career adjustment of Asian American workers (Leong, 2001). Although a great deal of research has been conducted concerning acculturation (see sources cited above), there is a paucity of research into the effect of acculturation on human performance as this is usually studied by HF/E researchers.

Ryder et al. (2000) have conducted research indicating that there are at least two dimensions to acculturation, for someone whose heritage (i.e., ancestral) culture is not the same as the mainstream culture. One dimension involves degree of identification one has with the mainstream culture in which one is embedded. Another dimension involves the degree of identification one has with one's heritage culture. (For multiethnic individuals, there would be additional dimensions of heritage culture identification, one each for each heritage culture; multiethnic individuals, of course, are increasing in number in our society, as noted by Root, 1996.) Ryder et al.'s research indicates that these two dimensions are orthogonal. The implication for HF/E research is that one must take at least a bidimensional approach to studying the effect of acculturation on performance, and assess identification with both mainstream and heritage cultures.

DISCUSSION

There is a whole spectrum of research areas in which we should expect worldview and acculturation to exert significant effects, one of which is *team cognition*. Aside from the choice of a topical area, *assessment methodology* is crucial for any research effort. We discuss these two issues below.

Team Cognition

It has been posited that culture has a strong influence on role conceptions (particularly role expectations) and role behaviors in teams and organizations (Stone-Romero, Stone, & Salas, 2003). This raises the issue of the effect of worldview and acculturation on team cognition.

Team cognition has been a concern in performance studies for years (Cannon-Bowers & Salas, 2001). It has been asserted that one area of agreement across various approaches to team cognition is "the fundamental hypothesis that common cognitions among team members will be associated with team effectiveness" (Rentsch & Woehr, 2004, p. 11). There is evidence for

this proposition in relation to shared mental models (e.g., Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). It would be worthwhile to test this hypothesis in relation to shared (or unshared) worldview and acculturation, as aspects of team cognition.

The effect of worldview and acculturation similarities and differences among team members on team performance is a matter of concern to some researchers. Such issues have been raised with regard to sports teams (Kontos & Breland-Noble, 2002), and it would be useful to investigate these issues with regard to many different types of teams.

Although a meta-analysis of research indicated no main effect for *personality* homogeneity among work team members (Bowers, Pharmer, & Salas, 2000), it must be noted that neither worldview nor acculturation were considered in the studies surveyed. Worldview and acculturation are more properly considered as cognitive constructs than as personality constructs; hence, the applicability of personality research to the domain of worldview and acculturation is debatable. Consequently, the effect of worldview and acculturation homogeneity on team performance is still an open question.

We suspect that worldview/acculturation *complementarity* (i.e., 'cover-all-bases' heterogeneity) may be at least as important as homogeneity, at least for certain kinds of tasks. That is, some group tasks may require a certain degree of worldview/acculturation *difference* among group members in order to be addressed optimally; this is suggested by Triandis, Hall, and Ewen's work (1965). Other types of tasks may require worldview/acculturation homogeneity. Thus, using the people-, data-, and thing-oriented taxonomy of task information posited by Hogan, Broach, and Salas (1990), perhaps there are different optimal worldview/ acculturation homogeneity/heterogeneity profiles for tasks in different parts of the taxonomy.

Assessment Methods

Several instruments are available for the assessment of various dimensions of worldview. Perhaps the oldest such instrument is Rotter's (1966) I-E scale, to assess locus of control. The Worldview Assessment Instrument (Koltko-Rivera, 2000) assesses mutability, relation to group (I-C), relation to authority, and locus of responsibility (related to locus of control). The Scale to Assess World Views (Ibrahim & Kahn, 1987) assesses time orientation, among other dimensions. It should be noted that the development of instrumentation to assess worldview dimensions is a pressing need.

Many instruments have been developed to assess acculturation (see review in Zane & Mak, 2003). Heritage culture acculturation usually is measured separately for different heritage cultures. That is, there are tests for different types of Hispanic acculturation, Asian acculturation, African American racial identity, and so on. (In passing, we note that it would be useful to have a single "heritage culture acculturation" scale, although this raises some thorny theoretical issues.)

CONCLUSION

The worldview and acculturation constructs are highly relevant to the concerns of HF/E researchers. We hope researchers will include these classes of variables within HF/E studies, including cutting-edge projects. (For example, it would be particularly interesting to research the way in which the performance correlates of worldview and acculturation are mediated by the telepresence variable, given that performance measures for distributed environments are now available; Dwyer, Oser, Salas, & Fowlkes, 1999.) We believe that the effort involved in adding these variables to the research protocol will be rewarded in terms of added performance outcome variance explained.

ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported in part by the U.S. Department of Defense under grant DAMD17-02-C-0016 (Principal Investigator: M. E. Koltko-Rivera), and by the U.S. Department of Defense MURI program administered by the Army Research Office under grant DAAD19-01-1-0621 (Principal Investigator: P. A. Hancock). E-mail regarding this presentation may be addressed to the first author at <koltkorivera@yahoo.com>.

REFERENCES

Arbuckle, T. Y., Gold, D. P., Andres, D., Schwartzman, A., & Chaikelson, J. (1992). The role of psychosocial context, age, and intelligence in memory performance of older men. *Psychology & Aging, 7*, 25-36.

Berry, J. W., & Sam, D. L. (1997). Acculturation and adaptation. In J. W. Berry, M. H. Segall, & C. Kagitçibasi (Eds.), *Handbook of cross-cultural psychology: Vol. 3. Social behavior and applications* (2nd ed., pp. 291-326). Boston: Allyn & Bacon.

Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin, 119*, 111-137.

Bowers, C. A., Pharmer, J. A., & Salas, E. (2000). When member homogeneity is needed in work teams: A meta-analysis. *Small Group Research, 31*, 305-327.

Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior, 22*, 195-202.

Center for Social Epidemiology. (n.d.). Locus of control and cardiovascular health. Retrieved June 8, 2004 from Job Stress Network website: http://www.workhealth.org/risk/ rfblocus.html

Chun, K. M., Balls Organista, P., & Marín, G. (Eds.). (2003). *Acculturation: Advances in theory, measurement, and applied research*. Washington, DC: American Psychological Association.

Driskell, J. E., Salas, E., & Johnston, J. (2000). Does stress lead to a loss of team perspective? *Human Performance in Extreme Environments, 5*, 69-76.

Dwyer, D. J., Oser, R. L., Salas, E., & Fowlkes, J. E. (1999). Performance measurement in distributed environments: Initial results and implications in training. *Military Psychology, 11*, 189-215.

Erez, A., & Judge, T. A. (2001). Relationship of core self-evaluations to goal setting, motivation, and performance. *Journal of Applied Psychology, 86*, 1270-1279.

Fitts, P. M. (1963). Engineering psychology. In S. Koch (Ed.), *Psychology: A study of a science: Vol. 5* (pp. 908-933). New York: McGraw-Hill.

Flores, E., Tschann, J. M., Marin, B. V., & Pantoja, P. (2004). Marital conflict and acculturation among Mexican American husbands and wives. *Cultural Diversity & Ethnic Minority Psychology, 10*, 39-52.

Hancock, P. A. (Ed.). (1999). *Human performance and ergonomics*. San Diego, CA: Academic.

Helander, M. G. (1997). The human factors profession. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 3-16). New York: Wiley.

Hockey, G. R. J., & Hamilton, P. (1983). The cognitive patterning of stress states. In G. R. J. Hockey (Ed.), *Stress and fatigue in human performance* (pp. 331-362). New York: Wiley.

Hogan, J., Broach, D., & Salas, E. (1990). Development of a task information taxonomy for human performance systems. *Military Psychology, 2*, 1-19.

Ibrahim, F. A., & Kahn, H. (1987). Assessment of world views. *Psychological Reports, 60*, 163-176.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluation traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86*, 80-92.

Kagitçibasi, C. (1997). Individualism and collectivism. In J. W. Berry, M. H. Segall, & C. Kagitçibasi (Eds.), *Handbook of cross-cultural psychology: Vol. 3. Social behavior and applications* (2nd ed., pp. 1-49). Boston: Allyn & Bacon.

Kaslow, N. J., Price, A. W., Wyckoff, S., Grall, M. B., Sherry, A., Young, S., et al. (2004). Person factors associated with suicidal behavior among African American women and men. *Cultural Diversity & Ethnic Minority Psychology, 10*, 5-22.

Kim, I. J., & Zane, N. W. S. (2004). Ethnic and cultural variations in anger regulation and attachment patterns among Korean American and European American male batterers. *Cultural Diversity & Ethnic Minority Psychology, 10*, 151-168.

Koltko-Rivera, M. E. (2000). The Worldview Assessment Instrument (WAI; doctoral dissertation, New York University, 2000). *Dissertation Abstracts International, 61* (04), 2266B. (UMI Microform No. 9968433)

Koltko-Rivera, M. E. (2004). The psychology of worldviews. *Review of General Psychology, 8*, 3-58.

Koltko-Rivera, M. E., Ganey, H. C. N., Dalton, J., & Hancock, P.A. (2004). *Human performance: A comprehensive meta-theory*. Manuscript in preparation, University of Central Florida.

Kontos, A. P., & Breland-Noble, A. M. (2002). Racial/ethnic diversity in applied sport psychology: A multicultural introduction to working with athletes of color. *Sport Psychologist, 16*, 296-315.

Lefcourt, H. M. (1982). *Locus of control: Current trends in theory and research* (2nd ed.). Hillsdale, NJ: Erlbaum.

Leong, F. T. L. (2001). The role of acculturation in the career adjustment of Asian American workers: A test of Leong and Chou's (1994) formulations. *Cultural Diversity & Ethnic Minority Psychology, 7*, 262-273.

Liem, R., Lim, B.A., & Liem, J. H. (2000). Acculturation and emotion among Asian Americans. *Cultural Diversity & Ethnic Minority Psychology, 6*, 13-31.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*, 273-283.

Phillips, J. M., & Gully, S. M. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology, 82*, 792-802.

Rentsch, J. R., & Woehr, D. J. (2004). Quantifying congruence in cognition: Social relations modeling and team member schema similarity. In E. Salas & S. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 11-31). Washington, DC: American Psychological Association.

Root, M. P. P. (1996). The multiracial experience: Racial borders as a significant frontier in race relations. In M. P. P. Root (Ed.), *The multiracial experience* (pp. xiii-xxvii). Thousand Oaks, CA: Sage.

Rotter, J. B. (1966). Generalized expectations for internal versus external control of reinforcement. *Psychological Monographs, 80*(609), 1-28.

Ryder, A. G., Alden, L. E., & Paulhus, D. L. (2000). Is acculturation unidimensional or bidimensional? A head-to-head comparison in the prediction of personality, self-identity, and adjustment. *Journal of Personality and Social Psychology, 79*, 49-65.

Smith, P. B., & Schwartz, S. H. (1997). Values. In J. W. Berry, M. H. Segall, & C. Kagitçibasi (Eds.), *Handbook of cross-cultural psychology: Vol. 3. Social behavior and applications* (2nd ed., pp. 77-118). Boston: Allyn & Bacon.

Stone-Romero, E. F., Stone, D. L., & Salas, E. (2003). The influence of culture on role conceptions and role behavior in organisations. *Applied Psychology: An International Review, 52*, 328-362.

Technical Group on Organizational Design and Management, Human Factors Society. (1984). In H. W. Hendrick & O. Brown, Jr. (Eds.), *Human factors in organizational design and management* (p. vii). Amsterdam: North-Holland.

Triandis, H. C., Hall, E. R., & Ewen, R. B. (1965). Member heterogeneity and dyadic creativity. *Human Relations, 18*(1), 33-35.

Wheeler, L., Reis, H. T., & Bond, M. H. (1989). Collectivism-individualism in everyday social life: The Middle Kingdom and the melting pot. *Journal of Personality & Social Psychology, 57*, 79-86.

Zane, N., & Mak, W. (2003). Major approaches to the measurement of acculturation among ethnic minority populations: A content analysis and an alternative empirical strategy. In K. M. Chun, P. Balls Organista, & G. Marín (Eds.), *Acculturation* (pp. 39-60). Washington, DC: American Psychological Association.

Zoroya, G. (2003, October 13). Army probes soldier suicides. *USA Today*, p. 1A.

THEOR. ISSUES IN ERGON. SCI.
NOVEMBER–DECEMBER 2004, VOL. 5, NO. 6, 499–506

Taylor & Francis
Taylor & Francis Group

COMMENTARY

On the relevance of qualitative methods for ergonomics

P. A. HANCOCK and J. L. SZALMA*

Department of Psychology and Institute for Simulation and Training,
University of Central Florida

Keywords: Qualitative methods; ergonomics; ethics; human factors; intention; science of purpose.

The application of qualitative methods to ergonomics research and practice offers us a new window on the nature of the interaction between humans and technology. The method discussed by Hignett and Wilson (2004) exemplifies this potential by applying their method to explicate the attitudes of practitioners and academic researchers toward qualitative methodology itself. Their specific findings, however, may be due in part to differences between the institutional structures in which they work as well as the attitudes of the specific individuals surveyed. Here, we offer a commentary on their work and reinforce the importance of qualitative research in ergonomics, while discussing the philosophical empirical, and theoretical issues raised by the introduction of these methods. We conclude that the fundamental problems inherent in qualitative approaches are limitations on quantitative methods also, being inherent to all forms of observation. While supportive of the general thesis proposed and especially appreciative of the authors' focus on purpose, we point to the problem of integrating different orders of knowledge as a significant barrier to future progress towards a comprehensive theory for ergonomics.

1. Introduction

There will always be worries and concerns when we are asked to move beyond our tried and tested approaches and embrace something that has the appearance of novelty. In any profession or academic enterprise there will be those who hold hard to the traditional and disdain the new and, conversely, there will be a group that embraces the new solely because of its innovative character. This leaves the preponderance of individuals cautiously interested but, as yet, unpersuaded. In trying to persuade the undecided, what Hignett and Wilson (2004) offer is the opportunity to embrace a wider vista of investigative procedures but one that is fraught with perceived and real peril since it lies at the very edge of science itself. About a century ago, psychology faced the exact reverse of this problem. Replete with subjectivity, the area seemed primed to divorce itself from all objective observation. Then, it was only through Watson's (1913) draconian dissociation from all inner experience that psychology fought to re-enter the realms of 'science.' Today,

*Author for correspondence. e-mail: jszalma@pegasus.cc.ucf.edu

Theoretical Issues in Ergonomics Science ISSN 1463–922X print/ISSN 1464–536X online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI 10.1080/1463922041233130339 1

we in ergonomics should be wary of any absolutes with respect to research procedures and should be more embracing of diverse ways of knowing. Indeed, in dealing with human beings, ergonomics is categorically unable to intrinsically limit itself solely to mathematical and quantitative methodologies. Like ethnology and anthropology, ergonomics must seek diverse strategies to comprehend the complexities of behavior in context. In throwing away the blinders of so-called 'physics envy,' let us now evaluate some such methods which Hignett and Wilson (2004) set before us.

2. Philosophical issues

Hignett and Wilson (2004) argue that the personal philosophy of the practitioner plays a critical role in the choice of method of assessment and approach to ergonomics. We agree with this assertion, as it points to the primacy of intention and purpose of behavior (Hancock 1995). However, we would point out that a sound philosophical position is important in any scientific endeavor, regardless of whether it is qualitative or quantitative in nature.

Hignett and Wilson bolster their argument for using qualitative methodology by appealing to transcendental realism, a position that allows for the existence of physical structures outside the mind but holds that different persons can have distinct but legitimate perceptions of that external world. However, qualitative methodology suffers the same limitation as quantitative methods under this philosophy, because the fundamental problem remains unresolved. That is, how does one establish the 'true condition' (analogous to the 'true score' in classical measurement theory)? Further, given the 'reality' of multiple realities, can such a 'true' condition be said to exist at all? Allowing for multiple realities does not solve this problem because the issue of the existence of an objective universe (i.e. a universe independent of perception and cognition) is not addressed. We do not expect Hignett and Wilson's qualitative methods to address this complicated issue, much less solve such a weighty philosophical conundrum. Nor do we believe that such unresolved problems with their philosophical bases for qualitative research necessarily reduces the immediate utility of the methodology or its relevance. Indeed, we praise them for bringing attention to this kind of research in a field that has historically been steeped in quantitative-based experimental research and expressed largely within an engineering framework. We do however suggest that the philosophical foundation for their efforts needs clarification and strengthening as the hoped for efforts to integrate qualitative findings increase.

3. Empirical issues

Ultimately the question of qualitative versus quantitative methods misses the fundamental point, which is the same for both: how can we verify our empirical evidence? This is an old question, regarding how we know that we are asking the correct questions of nature and the criterion for accepting evidence adduced. In other words, given that different people can have different perceptions of the same event, what constitutes valid empirical evidence? It is a problem that one cannot avoid by using any particular methodology, as the issue remains whether one measures physical parameters, physiological indices, performance response, subjective apperceptions, or qualitative interview data. Far from weakening the argu-

ment for adopting qualitative methods, this ambiguity strengthens it. If we cannot know with certainty the veracity of our empirical evidence, then assessing nature in multiple ways provides a better picture than a catholic devotion to a single methodology. With all due respect to Lord Kelvin[1], we argue that empirical evidence does not necessarily have to be quantitative. Any object or event, being constrained by space and time, has at least one characteristic that can be measured quantitatively, but this does not mean that quantitative data are the *only* way to characterize such an object or event. It will have characteristics that can be described qualitatively as well. Hence, a mature investigative philosophy should embrace both forms of assessment. Using the two methods conjunctively can yield a deeper understanding. This is especially true of human behavior and mental processes since they contain esoteric components, which stand in direct contrast to the traditional engineering concerns that deal with inanimate material. While qualitative methods are currently used informally in quantitative experimental research (e.g. in the form of debriefing at the end of experimental sessions), we should, as Hignett and Wilson indicate, consider adopting more formal qualitative procedures to enhance our understanding of human behavior.

In describing qualitative methods, Hignett and Wilson explicitly acknowledge that such research is subjective. Such subjectivity is disturbing to many ergonomists who, eager to reside under the engineering umbrella, are loathe to use methods that may appear 'unscientific' (see also Hancock *et al*. 2002). However, quantitative and 'objective' methods are each hardly independent of the experimenter. The subjectivity of the researcher is an issue regardless of the research method. The care taken to address the issue in qualitative research should therefore be transferred to quantitative research as well. It is interesting that psychologists, qua scientists, profess such an aversion to qualitative methods (Hancock *et al*. 2002). Other disciplines do not express such disdain, as is evident in many of the biological sciences. Chemistry also involves qualitative assessment. For instance, reports of syntheses of new molecules are usually accompanied by descriptions of qualitative characteristics such as color, smell (in the older literature, before the advent of safety procedures), and appearance (e.g. liquid, solid, crystalline, etc). In inorganic chemistry, metals are categorized as 'hard' or 'soft.' If biologists and chemists use descriptive methods when appropriate, why should we shirk from applying these useful methods to psychology and ergonomics? It is not less scientific to do so.

4. Theoretical issues: the importance of purpose

Hignett and Wilson (2004) see ergonomics as a socially situated practice. The social context makes the issue of intention critical as part of theory development and we wholeheartedly agree with the authors' contention that ergonomics must extend its view to include social and emotional factors to address the 'why' in design and behavior generally (see Hancock 2000, Murphy *et al*. 2003). Can qualitative methods aid in revealing intention and describing not only 'what is' but also 'what should

[1] William Thomson (Lord Kelvin), who devised the Kelvin (ratio) scale of temperature, asserted that 'when you measure what you are speaking about and express it in numbers, you know something about it, but when you cannot express it in numbers your knowledge about it is of a meager and unsatisfactory kind (D MacHale, Comic Sections)'.

be'? (Hancock and Diaz 2002). If we accept that purpose is essential to ergonomics and sets it apart from other theories of behavior in context (Hancock 1997, Hancock and Diaz 2002), and that ethics is important for ergonomic theory (Shipley, 1998), then promoting the development of such a unique theory requires aggregation of seemingly disparate methods from vastly different disciplines. Qualitative methodology offers one innovative and critical component of this aggregation.

It is possible that using qualitative methods can enhance the development of ergonomic theory promoted by Hancock and Diaz (2002). As an applied field, ergonomics has, at best, generated microtheories, and in many applications empirical work is atheoretical. While Newton's assertion, '*hypothesis non fingo*,' (meaning that description is sufficient to itself and causation is not promulgated) has merit, we believe that theory building and theory testing is critical to ergonomic advancement at this time. By providing a different window to the world, this form of empirical method can provide a unique dataset that permits creation of comprehensive theory of what the human-technology interface *should* be. It can also reveal user intention in a way that can inform the development of design principles. Indeed, if a comprehensive ergonomic theory lies in a theory of intention grounded in linking ecological and information processing views, as suggested by Hancock and Diaz (2002), then qualitative methods have much to offer in specifying the perception of environmental characteristics and quality of information processing. It may help in developing what Hancock and Diaz (2002) called 'mindful technology' rather than the mindless technology that currently is propagated in large volume in the industrialized world.

Hignett and Wilson argued that qualitative methods will take ergonomics to level of their table 1 (spiritual sciences) in which 'universal purpose' is adopted as a central concept. This accords with the arguments of Hancock and Diaz (2002) for ergonomics providing the cornerstone to a science of purpose. It is argued here that qualitative methods are a useful tool for beginning to build this edifice. It is not our intention to challenge the usefulness of quantitative research. Rather, we see the two as complements rather than antagonists, in that they provide different vistas into phenomena of interest. Each provides data on an aspect or set of characteristics of the studied object, and like a gestalt, perhaps a complete picture of the whole only comes from embracing multiple measurement methods. Given the importance of both methods, it may not be advisable to divide different disciplines into qualitative and quantitative, as Hignett and Wilson have done in their figure 2 (e.g. social psychology in the qualitative end and experimental psychology at the quantitative end). While these distinctions may be historically correct, both methods could, in principle, be used by any of the human sciences.

5. Academics versus practitioners

An important question raised by Hignett and Wilson's work is not whether qualitative methods should be employed by ergonomics, but for what applications they should be used. It is not a question of quantitative versus qualitative, but of when a particular tool is appropriate. It is easy to allow this distinction to fall along academic/practitioner lines, but this imposes a categorization that most likely does not describe the many ergonomics professionals who straddle both domains. Some might argue that qualitative methods are more appropriate for the practitioner, for whom case study methods are useful in diagnosing and solving practical problems. However, qualitative methods have the potential to enhance the efforts of academic

researchers as well as practitioners. For instance, a detailed qualitative study of domain experts (e.g. professional drivers, industrial workers, military personnel) can inform the design of more traditional quantitative laboratory studies, thereby enhancing the external validity of such research. Qualitative studies in tandem with the quantitative can thus help address a long standing problem of external validity in basic laboratory research by aiding in the design of ecologically valid laboratory displays and controls. To date, qualitative research in experimental psychology has largely restricted itself to information gathered during the debriefing of participants at the end of an experiment (but see Hutchins 1996). This lack of systematic attention paid to collecting qualitative data is at odds with the tradition of rigorous application of quantitative methodology by experimental psychologists in the design, execution, and analysis of their experiments. It can likely be traced to the behaviorist disdain for studying mental events and the general desire for psychology to be quantitative based on a misguided belief that 'real science' is fundamentally mathematical in nature.

Hignett and Wilson noted that academics and practitioners have different views on the gap between the internal and external environments of ergonomics, in which academics see themselves as filling the gap while practitioners see the gap as reflecting differing views. Each author has a different explanation for why academics and practitioners differ in their viewpoints. One author suggests that it is due to a difference in approach, with academics taking a more traditional quantitative approach, while the other believes that academics feel they need to be jacks-of-all trades. We suggest here that the distinction is due to the behaviorist tradition in academia and the pragmatic considerations of practitioners. To use an analogy, academics and practitioners work in different environments with different affordances and with different goals (intention) which shape their behavior in different directions.

6. Unity and diversity

There was a time when those with a thirst for understanding did not divide the world into component parts for study. All was seen as facets of a unified whole. However, as the world became more complex, logical ways were sought to partition inquiry and derived disciplines subsequently focused on discrete aspects of the world. Art contrasted with science, subjective with objective perspectives and, potentially worst of all, teleology was divorced from technology (see Hancock 1997). This 'divide and study' strategy served well to advance discrete understanding, but now we are realizing that the next step is in integration and cross-fertilization of these divisions in the search for new and integrative knowledge. This coming together or 'consilience' as E. O. Wilson (1998) terms it, is now evident in the world of ergonomics, which stands at the confluence of art and science, of design and technology, and, as Hignett and Wilson (2004) elucidate, qualitative and the quantitative forms of understanding. This process of vergence and emergence is not easily accomplished. It is tantamount to mixing the oil of engineering with the cerebral-spinal fluid of neuroscience. There are many barriers to such a meeting of minds, not the least of which is academic tradition. In general, ergonomics has been a discipline without a home in academic institutions, which are divided traditionally into disciplines represented by distinct 'departments.' Practitioners often evolve from the confluence effects we have noted above, but there are few analogs in the academic world. The structure of

academic institutions, as currently configured, does not provide a receptive scaffolding on which ergonomics can be supported. Indeed, we argue that despite the polemics of many administrators, the traditional discipline-based divisions of academic institutions act to discourage inter-disciplinary work which is at the heart of ergonomic science. We suggest that the results that Hignett and Wilson (2004) report in regard to practitioner vs. academic attitudes may, in part, reflect these structural institutional differences.

7. Conclusion

Hignett and Wilson (2004) describe a useful and innovative approach for analyzing data from human-machine systems research. As such, it is an excellent addition to the toolbox of both researchers and practitioners. Its application to the issue of attitudes of academics and practitioners toward these methods, however, misses important differences between the contexts in which these two groups work. This underscores the importance of developing qualitative research questions in an iterative manner, as well as the necessity to consider the context in which the behaviors of interest evolve. In sum, Hignett and Wilson (2004) are asking us to learn new languages and embrace wider vistas. They have disturbed the comfortable, but we are comfortable with having been disturbed.

Acknowledgements

This research was facilitated by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, Dr. Elmar Schmeisser, Technical Monitor (Grant # DAAD19-01-1-0621). The research was also facilitated by grants from the Army Research Laboratory, Mr. John Lockett, Dr. Michael Barnes, and Dr. Jessie Chen, Technical Monitors (Grant # DAAD 19-01-C-0065), and from a DARPA-funded program under Grant NBCH1030012, CMDR Dylan Schmorrow, Technical Monitor. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government.

References

Dempsey, P. G, Wogalter, M. S. and Hancock, P. A. 2000, What's in a name? Using terms from definitions to examine the fundamental foundation of human factors and ergonomics science, *Theoretical Issues in Ergonomic Science*, **1**(3), 3–10.

Hancock, P. A. 1995, Teleology for technology, in R. Parasuraman and M. Mouloua (eds), *Automation and Human Performance: Theory and Applications* (Hillsdale, NJ: Erlbaum), 461–497.

Hancock, P. A. 1997, Men without machines, in P. A. Hancock (ed.), *Essays on the Future of Human-Machine Systems* (Bunta: Eden Prairie, MN), 137–142.

Hancock, P. A. 2000, A stranger in paradigms. Keynote Paper. In *Proceedings of the 14th Triennial Congress of the International Ergonomics Association and the 44th Annual Meeting of The Human Factors and Ergonomics Society*, San Diego, California, USA. July 29–August 4, 2000. (Santa Monica: The Human Factors and Ergonomics Society).

Hancock, P. A. and Diaz, D. D. 2002, Ergonomics as a foundation for a science of purpose. *Theoretical Issues in Ergonomic Science*, **3**, 115–123.

HANCOCK, P. A., WEAVER, J. L. and PARASURAMAN, R. 2002, Sans Subjectivity – Ergonomics is Engineering, *Ergonomics*, **45**, 991–994.

HIGNETT, S. and WILSON, J. R. 2004, The role for qualitative methodology in ergonomics: a case study to explore theoretical issues. *Theoretical Issues in Ergonomics Science*, **5**, 473–493.

HUTCHINS, E. 1996, *Cognition in the wild* (Cambridge, MA: MIT Press).

MURPHY, L., STANNEY, K. and HANCOCK, P. A. 2003, The effect of affect: the hedonomic evaluation of human–computer interaction, *Proceedings of the Human Factors and Ergonomics Society*, **47**, 764–768.

SHIPLEY, P. 1998, The ethical turn and the workplace. The Ergonomics Society Lecture, *Ergonomics*, **41**(1), 1–19.

WATSON, J. 1913, Psychology as a behaviorist views it, *Psychological Review*, **20**, 158–177.

WILSON, E. O. 1998, *Consilience the unity of knowledge* (New York: Knopf).

About the authors

Peter A. Hancock is Provost Distinguished Research Professor in the Department of Psychology, the Institute for Simulation and Training, and at the Department of Civil and Environmental Engineering at the University of Central Florida. He is currently the Director of the Minds in Technology, Machines in Thought Laboratory at UCF. In his previous appointment, he founded and was the Director of the Human Factors Research Laboratory at the University of Minnesota. At Minnesota he held appointments as Full Professor in the Departments of Computer Science and Electrical Engineering, Mechanical Engineering, Psychology, and Kinesiology as well as at the Cognitive Science Center and the Center on Aging Research. He currently holds a courtesy appointment as a Research Scientist at the Massachusetts Institute of Technology (MIT) and as an Adjunct Senior Research Scientist at the Transportation Institute of the University of Michigan. Professor Hancock is the author of over four hundred refereed scientific articles and publications as well as editing numerous books including: Human Performance and Ergonomics in the Handbook of Perception and Cognition series, published by Academic Press in 1999 and Stress, Workload, and Fatigue, published in 2001 by Lawrence Erlbaum. He is the author of the 1997 book, Essays on the Future of Human-Machine Systems. He has been continuously funded by extramural sources for every year of his professional career, including support from NASA, NIH, NIA, FAA, FHWA, the US Navy and the US Army as well as numerous State and Industrial agencies. He is the Principal Investigator on the recently awarded Multi-Disciplinary University Research Initiative, in which he will oversee $5 Million of funded research on stress, workload, and performance. In 1999 he was the Arnold Small Lecturer of the Human Factors and Ergonomics Society and in 2000 he was awarded the Sir Frederic Bartlett Medal of the Ergonomics Society of Great Britain for lifetime achievement. He was the Keynote Speaker for the International Ergonomics Association and the Human Factors and Ergonomics Society at the 2000 combined meeting in San Diego. In 2001 he won the Franklin V. Taylor Award of the American Psychological Association as well as the Liberty Mutual Prize for Occupational Safety and Ergonomics from the International Ergonomics Association. In association with his colleagues Raja Parasuraman and Anthony Masalonis, he was the winner of the Jerome Hirsch Ely Award of the Human Factors and Ergonomics Society for 2001, the same year in which he was elected a Fellow of the International Ergonomics Association. He was awarded a Doctor of Science (D.Sc.) degree from Loughborough University in December, 2001. In 2002, he was awarded the Jastrzebowski Medal of the Polish Ergonomics Society for contributions to world ergonomics and in the same year was named a Fellow of the Ergonomics Society of Great Britain. He has recently been elected to a three-year term as a Member of the National Research Council's Committee on Human Factors and which will run concurrently with his membership of the Executive Council of the Human Factors and Ergonomic Society. His current experimental work concerns the evaluation of behavioral response to high-stress conditions. His theoretical works concerns human relations with technology and the possible futures of this symbiosis. He is a Fellow of and past President of the Human Factors and Ergonomics Society. He collects and studies antique maps and is a committed Ricardian.

James L. Szalma is a Senior Research Scientist in the Minds in Technology, Machines in Thought Laboratory at the University of Central Florida. His previous appointment was assistant professor in the psychology department at SUNY Farmingdale. He received his Ph.D. in 1999 from the University of Cincinnati, where he investigated the performance, workload, and stress of monitoring tasks, and the use of feedback in training for vigilance performance. In addition to vigilance, his current research efforts are aimed at testing theoretical models of stress and performance. As part of this work, he is investigating the individual differences related to stress and coping strategies. he is also conducting studies to empirically test a novel modification of Signal Detection Theory, Fuzzy Signal Detection Theory.

THE ADVERSE IMPACT OF REMOTE COMMAND AND CONTROL UNDER LIVE FIRE

Kip Smith
Linköping University
SE 581 83 Linköping, Sweden

The US Army is planning a shift in the mode of command and control between soldiers and their battlefield leaders. Soldiers will receive orders from afar through electronic means of communication. This practice is known as remote command and control. Reliance on remote command and control has the potential to erode trust between soldiers and their leaders and should, we hypothesize, slow soldiers' response times to commands to move and to shoot. We have conducted two field experiments to test this hypothesis (Pangburn, Freund, Pangburn, & Smith, 2003). Our laboratory was a paintball assault lane. While a paintball lane is not actual combat, we have found that its live but non-lethal fire makes it an effective and ethical laboratory for studying the behavioral and cognitive effects of stress induced by live fire. Participants were exposed to two conditions of communication mode, leader-present (face to face) and leader-remote (two-way radio). The contrast between participants' response times to commands to move and to shoot was statistically significant and was consistent with the predicted decrement in the remote command-and-control condition.

INTRODUCTION

The pervasive specter of death makes the battlefield a prototypical extreme environment. Troop survival requires vigilant and efficient teamwork, effective leadership, and the sporadic madness of heroism. The effectiveness of a battlefield leader is a direct function of the trust felt by the members of the fighting unit. Trust in the leader makes the battlefield feel less extreme, promotes teamwork, and encourages heroism.

The link between a battlefield unit and its immediate leader has traditionally been verbal and direct. Face-to-face contact and mutual hardship have long been the forge of trust. To use a term coined during the recent invasion of Iraq, trust emerges because the leader has been 'embedded' in the unit.

For soldiers in the U. S. infantry and their immediate leaders, this foundation of trust is about to disappear. The battlefield environment will soon become even more extreme.

The fundamental change is in the mode of communication between soldiers on the battlefield and their immediate leaders. One of the cornerstones of the U. S. Army's plan for modernizing the infantry is to have soldiers on the battlefield receive orders from afar through electronic means of communication. This practice is known as 'remote command and control.' The soldiers who will operate under this plan will no longer take their battle commands from a leader standing within visual range. The only connection with their commanding officers will be their radios or other electronic information devices.

We suspect that this fundamental organizational change is likely to do more than remove the leader from the battlefield. It will also erode the development of

common ground (Clark, 1996) between soldiers and their leader. We expect the erosion of common ground to have profoundly negative implications for the development of trust and for the effectiveness of the fighting unit.

Common ground is special form of intentional thinking (e.g., Dennett, 1997). You have an intentional thought whenever you think about another person's (or group's) thinking. You and another (person or group) have common ground whenever you both know what each other knows and you both know that each other knows it (Clark, 1996). This mutual knowledge of a shared representation of information is the essence of common ground. The military works troops hard during basic and more advanced training so that they may develop common ground through mutual hardship and interdependence. Common ground has traditionally been the glue that holds a fighting unit together.

We believe that common ground works to support the development of trust in a leader and in the leader's authority. A leader who is (and has always been) in the field with his troops should have more credibility than an absent or remote leader. Accordingly, the use of remote communication can be expected to degrade or eliminate the mutual knowledge of a shared representation of information and, as a result, the leader's ability to exercise authority. We hypothesize that the lack of a basis for common ground will be reflected in slower reaction times when commands are given remotely via electronic devices than when they are given face-to-face.

In the two experiments presented here, we test this hypothesis in an ecologically valid simulation of the battlefield environment, namely, a paintball assault lane.

METHOD

Participants

Forty-two college students, five women and 37 men, drawn primarily from an introductory psychology class, volunteered to spend a Saturday afternoon playing paintball on our simulated assault lane. The average and median age was 19 with a range of 18 to 26. All signed informed consent forms and liability releases. All participants were treated according to APA guidelines.

Design, measures, and task

We manipulated leader presence at two levels (present and remote) in a within-subjects design. We measured the participant's response time to the leader's commands to move and to shoot. We predict slower response times in the leader-remote condition. We expect that the effect of leader presence will not be strong under normal circumstances, but that it will emerge in the relatively extreme environment of simulated combat.

The simulated combat environment in this study is a paintball assault lane, Figure 1. The lane consisted of eight protective stations behind which the participants could hide. At the end of the lane was a fortified position where a sniper was positioned. The sniper's task was to shoot participants as they moved up the lane. The participants advanced through the lane one at a time. Each participant had two tasks. The first task was to move from station to station up the lane in response to the command to move. The second task was to shoot targets in response to the command to shoot. Statistical analysis used a within-subjects t test of the mean difference in response times between the leader-remote and leader-present conditions. The response times to commands to move and shoot were recorded by an observer using a stopwatch.

For the first experiment, the US Army provided access to the 25 ft x 200 ft building at Range 52, Fort Riley, Kansas, home of the US Army 1st Infantry and an active training center for artillery. We set up our paintball assault lane in this building. As shown in Figure 1, the lane consisted of a staggered series of eight 'stations' behind which the participant could take cover. At the far end of the lane, a 'sniper' hid behind a screen of netting and protective cover. As this experiment was conducted inside a building with many windows, the leader in the remote condition was able to be completely out of the participant's sight while still able to view the assault lane.

Twenty students participated in the first experiment. Attendance was limited because a one-way trip to Fort Riley took 45 minutes and required passing through a security gate and a variety of active firing range complexes. All told, the experiment took at least four hours of the participants' time. In the interest of time and efficiency, we moved the second experiment to an on-campus facility (Quinlan Gardens). The setup was exactly the same as the lane in Figure 1 with one exception. The lane was set up in a small field rather than in a building. Twenty-two students participated in the second experiment.

The change in setting from inside to outdoors made the leader-remote condition in Experiment 2 less remote. In the leader-remote condition, the leader was approximately 100 yards behind the lane but still visible if the participant chose to look. In both experiments, all communication in the remote condition was by two-way radio.

Materials

Participants were given one paintball marker (gun), one set of US Army fatigues, a set of elbow and kneepads, and a paintball face shield. In the leader-remote condition, participants were also given a two-way radio and instructions on its use.

Procedure

When participants arrived at the lane, a uniformed soldier presented them with a standard battle-dress uniform and asked them to put the uniform on. The participants were then introduced to their leader, an army officer. The leader briefed them using the official military Operations Order format and addressed them by their last names. They were then told to assemble in a staging area where they could hear the activity in the assault lane while they waited their turns. All of this was purposefully done to immerse the participant in the experiment and to heighten the sense of realism and their anxiety.

While waiting, participants were instructed on the safety and use of the paintball markers and read a briefing. Participants were sent down the lane individually. Whenever the participants took aim at the targets or moved between stations, they exposed themselves to the sniper's fire. Participants were instructed to attempt to shoot enemy targets without hitting friendly targets. No measures were made of firing accuracy because our hypothesis concerns the participant's reaction time to commands given by the leader and says nothing about their marksmanship. The shooting task was created only to give focus to the participant's activity and to enhance the feel of a combat environment.

A small container with five paintballs was placed at each of the eight stations. The 40 paintballs in the eight containers were the participant's only ammunition. The participant started at one end of the lane, shown by the X in Figure 1. At this station and all subsequent stations, the leader gave the participant the command "Fire" when he judged it to be safe to do so. The time

elapsed from the issue of the command to the first shot fired is the first dependent measure.

When the participants ran out of paintballs, they reported "Out of ammo" to the leader, who then gave the command "Move" when he judged it safe to do so. The participants had to move across the lane to the next station and immediately pick up its container of five paintballs. The time elapsed between the issue of the command to move and the time the participant's hand first touched the new supply of ammunition was the second dependent measure. When the participants finished loading, they reported "Loaded" to the leader who then started the cycle over again with the command "Fire." To measure reaction time, an observer would shadow the participant and record elapsed time with a stopwatch. Although we assume this method was accurate and effective, in the future we hope to record these data using biometric telemetry.

In the leader-present condition, the leader was one station behind the participant. In the leader-remote condition the only contact between the leader and the participant was by two-way radio.

Every participant ran through the lane twice, once with the leader physically present and giving orders, and once with the leader viewing the lane from a remote location and giving orders through a two-way radio. The order of conditions was counterbalanced across participants.

The study was intended to generate some anxiety so that the measures would more readily generalize to the battlefield. The major sources of stress were the fear of being shot and actually being hit by paintballs. The pain associated with being struck by a paintball is slight but sudden and intense. Protective gear minimizes the risk of injury. A registered nurse received remuneration for being on site throughout the experiment.

RESULTS

The purpose of this study was to test whether we can observe any effect of leader presence on command and control in a simulated combat environment. The participants were exposed to two conditions, leader-present (face-to-face) communication, and leader-remote (two-way radio) communication.

Experiment 1

Figure 2a is a histogram of the within-subject difference in response times to the command to move for the participants in Experiment 1. The difference is calculated by subtracting the average of the times observed in the leader-present condition from the averages of the times observed in the leader-remote condition. Of the 20 participants, 15 took longer to move in the leader-remote condition than in the leader-present condition. A within-subjects t test for these data, $t(19) = 2.958$, $p < .004$, indicates that response times to the command to move were significantly longer in the leader-remote condition.

Figure 2b is the histogram of the within-subjects difference in response times to the command to shoot. The data are, once again, calculated by subtracting the time observed in the leader-present condition from the time observed in the leader-remote condition. Of the 20 participants, 12 took longer to shoot in the leader-remote condition than in the leader-present condition. A within-subjects t test for these data, $t(19) = 2.317$, $p < .016$, indicates that response times to the command to shoot were significantly longer in the leader-remote condition.

Experiment 2

Figures 3a and 3b are histograms of the within-subjects difference in response times to the commands to move and to shoot, respectively, for the participants in the second experiment. The data are calculated in the same way as those shown in Figure 2. Of the 22 participants, 16 took longer to move in the leader-remote condition than in the leader-present condition. A within-subjects t test for the move data indicates that the difference is significant $t(21) = 2.798$, $p < .006$. Of the 22 participants, 15 took longer to shoot in the leader-remote condition than in the leader-present condition. A within-subjects t test for the fire data indicates that the difference is significant $t(21) = 2.211$, $p < .020$.

Given the similarity of the two experiments' results, it appears the subtle difference in the degree of remoteness of the leader across the two experiments did not have a significant impact on response times. The similarity also allows us to aggregate the data. The test on the composite move data is significant $t(41) = 4.122$, $p < .0001$. The test on the composite shoot data is also significant $t(41) = 3.218$, $p < .0013$. These results support our hypothesis. Participants were faster to react to the leader's commands when the leader was present than when the leader was remote.

DISCUSSION

The two experiments reported here are the first in a planned series of studies on leader presence and common ground. Common ground is the foundation for the development of trust not only between troops and their leader but also among individuals. The battlefield is an extreme environment that serves as a metaphor for (too much of) everyday life. Understanding the impact of the erosion of common ground on the battlefield may suggest constraints for the design of tools and information devices that facilitate the development of common ground and, eventually, trust. This is our long-term goal. In the short term, data acquisition is continuing in a paintball assault lane in Sweden. We are now using automated biometric telemetry to eliminate the possibility of observer error in the measurement of response times and to assess psychophysiological indicators of stress.

In praise of the paradigm

It is simply not possible to simulate a combat environment in a laboratory. Our desire for ecological validity forced us to work outside the normal laboratory setting. For the first experiment, our assault lane was set in a building on an active Army base. The lane in the second experiment was set up in a lightly wooded field on campus.

Embedding experimental controls in a simulated combat environment is a new concept in human factors research. To attain ecological validity, we immersed participants into a pseudo-military environment by providing military uniforms, speaking in military command style, and by keeping the age group in line with that of the military. This procedure gives us the ability to generalize the results of the experiments to real-world battlefield settings.

Admittedly, paintball is not actual combat and our assault land cannot be said to be a truly extreme environment. Unlike a battlefield, the paintball assault lane allows us to exercise experimental control. However, it retains viable levels of several battlefield stressors including interactive hostility and the unannounced onset of sharp pain. The presence of these and other stressors makes the paintball lane an excellent natural laboratory for investigating the extreme environment faced by battlefield soldiers.

In addition to its face validity, the paintball paradigm has two highly desirable features. First, the availability of commercial paintball facilities make it possible for others to replicate and expand upon these and future experiments. Second, because paintball is a recreational activity (for the young), it has an ethical neutrality that makes it acceptable (if not palatable) to most members of Institutional Review Boards. This in itself is a coup. Paintball opens research on battlefield stressors to the scientific community at large.

AUTHOR NOTE

The views expressed in this work are those of the author and do not necessarily reflect official Army policy. This work was supported by the DOD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Dr. Elmar Schmeisser is the project monitor. The credit for conceiving these experiments and for their success goes to 2nd Lt. Keith Pangburn.

REFERENCES

Clark, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Dennett, D. (1987). *The Intentional Stance*. Cambridge: MA: MIT Press.

Pangburn, K., Freund, S., Pangburn, H., & Smith, K. (2003). Beyond Milgram: An experimental study of leader presence. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*. Denver, Colorado.

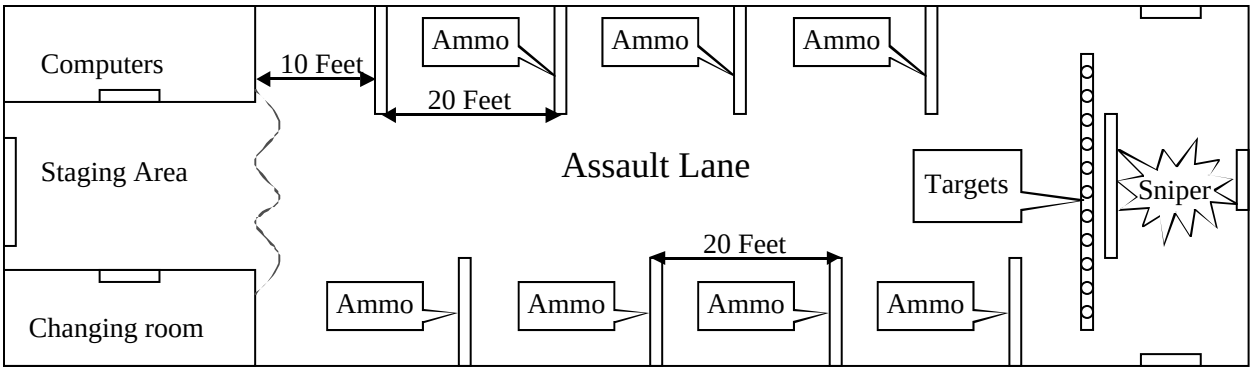


Figure 1 – The layout of the paintball assault lane

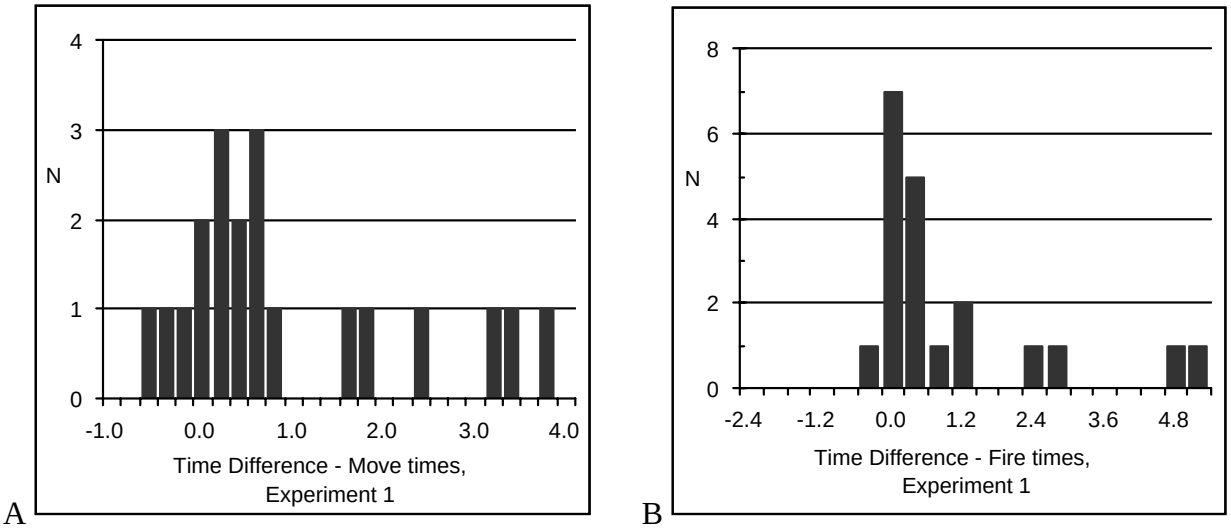


Figure 2 – (A) Histogram of the average differences in participants' response times to the command to move in Experiment 1. Data are calculated as the time to move from one station to the next in the leader-remote condition less the time in the leader-present condition. (B) Histogram of the average differences in participants' response times to the command to fire In Experiment 1.

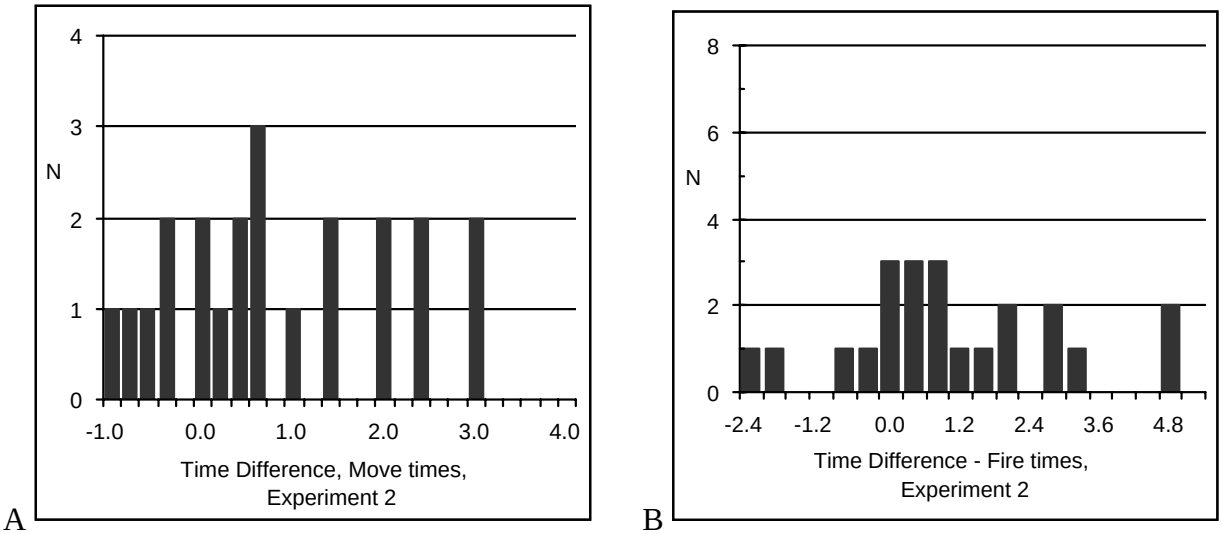


Figure 3 – (A) Histogram of the average differences in participants' response times to the command to move in Experiment 2. (B) Histogram of the average differences in participants' response times to the command to fire In Experiment 2.

Decision Support through Constraint Propagation in Collaborative Distributed Command and Control*

Rogier Woltjer
Graduate School of Human-Machine Interaction
Department of Mechanical Engineering, IKP/IAV
Linköping University, Sweden
rogwo@ikp.liu.se

Kip Smith
Department of Mechanical Engineering, IKP/IAV
Linköping University
Sweden
kipsm@ikp.liu.se

Abstract - *In this paper we develop a conceptual model of the interdependencies among plans that can be expected to emerge in a collaborative, distributed command and control center. The foundations of the model are the problem space representation of problem solving and analyses of the nature of constraints and their propagation and of the task of planning. The model has informed the development of a series of empirical studies of the propagation of constraints in a simulated command and control center. The C3Fire microworld is the test bed for the studies. Analysis of the behavioral data captured by C3Fire serves to inform the design of an intelligent interface for decision support in command and control that highlights constraints and facilitates human decision making.*

Keywords: Collaborative distributed decision making, decision support, constraints, intelligent systems, cognitive systems engineering.

1 Introduction

As Friman [1] and Brehmer and Sundin [2] indicate, decision making in the new vision of network-based command and control will be distinctly non-traditional. Staff will no longer relay orders down the hierarchy. In the network-based vision, the hierarchy is to be replaced by a 'command cell' of decision makers representing a broad spectrum of military and civilian competencies.

This paper focuses on the decision makers in the command cell. These agents have various competencies or constituencies that take part in the mission that is to be carried out (e.g. rescue workers, police, fire-fighters, medical personnel, air traffic controllers, etc.) They face the task of collaborating to establish a joint plan for their joint forces.

In the network-based vision, the decision makers will be situated in an environment where they make independent plans based on information that is only partially shared with the other agents. The decision makers are linked to their respective knowledge cell through personal information technology which enables them to conceive initial plans. They are linked together (to form the command cell) by innovative information technology that displays a shared representation of the agents' separate plans. The shared representation allows the separate plans to be merged and evaluated by all who are either physically or virtually present. The shared display should facilitate the tasks of (a) coordinating the separate plans, (b) assessing whether they are compatible, and (c) if they are not compatible, changing the plans so that they are compatible and can be executed.

We are developing a prototype of a decision-support system designed to help agents in a network-based command and control environment to detect the constraints on their actions and whether their separate plans are compatible. This work builds upon a variety of theoretical frameworks focusing on various characteristics of decision making: distributed decision making [3], dynamic decision making [4], and collaborative decision making [5], [6].

1.1 Problem solving and constraints

In their classic characterization of problem solving, Newell and Simon [7] define a problem as a gap between the solver's current state and goal state. For example, consider the problem faced by the fire chief awakened in the middle of the night by a jarring and largely incoherent telephone call that a fire is raging on the south ridge of Baldberg Peak. The fire chief's current state is one of fragmented uncertainty about a wild fire. If the chief is responsible for supervising fire-fighting efforts in the area, the goal state is likely to be the complete suppression of the fire. The gap is the difference between the current state of uncertainty and the goal state of putting the fire out. The fire chief's problem is posed by that gap. Problem solving is the course of action taken to close it.

Continuing with our paraphrasing of Newell and Simon's analysis, to close the gap is to solve the problem.

* **0-7803-8566-7/04/$20.00 © 2004 IEEE.**

There are two parts to this process. The first is to define the 'problem space' between the initial state and the goal state. The problem space is a network of intermediate states linked by incremental change in state. For example, among the nearest intermediate states for the newly awakened fire chief are brewing a pot of coffee and trying to remember whom to call for additional, more certain information. Focusing attention is a useful incremental step away from the initial state of almost any problem.

Hidden within the network of links and states in the problem space are paths from the initial state to the goal state. The second part of problem solving process is to take one of those paths, preferably a short one. Nowhere do Newell and Simon claim that these parts are sequential, that is, that problem solver must enumerate the galaxy of intermediate states in the problem space prior to taking steps to solve the problem. In practice, the problem space is defined by the constraints that emerge and propagate as the problem solver steps from state to state.

Each problem state, including the initial and goal states, has components that can be characterized as a set of constraints [8]. In our example, the initial state contains a geographic constraint (Baldberg Peak) and a temporal constraint (the middle of the night). These components of the initial state are exemplars of the class of constraints that Reitman calls 'closed' – they define, circumscribe, or otherwise limit the choice of paths through the problem space. The geographic constraint defines for the fire chief where attention needs to be focused, at least initially. The time constraint reduces the likelihood that reliable information can be found quickly and limits the readiness of first responders.

In addition, these closed constraints are likely to interact with the fire chief's domain knowledge to spawn additional constraints. Which response team or observer is closest to Baldberg Peak? Can they be relied upon to get to work in the middle of the night? The act of asking such questions is itself a transformation of the fire chief's current state. Answering them and acting upon those answers imposes additional closed constraints. Soon, the proliferation and propagation of constraints leads the problem solver down a path through the problem space.

Unfortunately, most paths through the problem space do not lead the solver to the goal state. For example, the problem solver who obtains inaccurate or biased information steps from one state to another but is unlikely to be headed directly towards the goal. Some paths loop back to the initial state. Others are 'garden paths' that appear enticing but that actually consume energy, time, and other resources as they lead the problem solver away from the goal. Garden paths and loops are most often taken when a set of closed constraints restricts the available options to only one or a very few intermediate states within the problem space. When guided solely by closed constraints the problem solver is likely to lose sight of the goal or control of the process (the suppression of the fire on Baldberg Peak).

The key to finding one of the paths that leads towards the goal state is the class of constraints that Reitman calls 'open.' Open constraints "provide definitional slack" [8, p. 293] by eliminating or redefining a component of the problem. Consider the fire chief who is mulling the question of whom to call for additional information or to mobilize to the site of the fire. The geographic constraint serves to close the options to the nearest available crew. But what to do if that crew has been overworked and deserves several days off? The answer is to open the geographic constraint by extending the scope of the search for a crew. By opening the previously closed constraint, the fire chief is likely to find a satisfactory alternative that leads to a different intermediate state and, hopefully, to a path to the goal state. The interaction of open and closed constraints defines the path through the problem space. Closed constraints shrink the problem space and reduce the degrees of freedom for taking a path through it. Open constraints produce flexibility and make paths available.

It is often the initial step that is the most difficult. The major source of difficulty is that almost every constraint is open. The problem solver is free to step into the problem space in any direction; there is an unbounded number of accessible intermediate states. Accordingly, problem solving typically begins by taking steps that close constraints. The closing continues to a point where a number of paths can be pursued and inspected. If the goal state can be seen down one of the paths, that path is taken and problem is solved (barring unexpected contingencies). If the goal cannot be seen, one or more of the closed constraints are opened, new states explored, and new paths taken.

1.2 Constraints in distributed collaborative decision making

Distributed decision makers necessarily formulate separate, local plans. They then propagate the constraints on their local plans to come to a common, global plan. Hutchins [9] provides an illustrative example, he observed operations on the bridge of a US Navy ship when an important navigation device had broken down. The ship's navigation team maintained control of the ship through self-organization of their problem solving and organizational structure. The team did not have a global problem description, but were able to come to a satisfactory global solution by setting constraints on each other's actions through presentation of partial solutions and solving problems locally. Hutchins showed that a "system of mutually adaptive computational sub-parts" [9,

p. 35] can solve problems efficiently in organizationally unanticipated situations by propagating mutual constraints.

When teams of decision makers are put together to solve a problem, finding these constraints can be difficult, because the decision makers are not familiar with the situations that they are forced to deal with. This paper hypothesizes that decision support tools can facilitate this process by enabling the decision makers to explicitly communicate the constraints they put on each other's actions. A salient characteristic of distributed or network-based decision making is that the environment is highly dynamic. Events and situations change constantly in the real world when extinguishing fires, rescuing victims of disaster, or fighting a battle. In a network architecture of command and control, the organizational structure must flex to reflect the changing set of agents who are best able to make decisions, perform actions, and take control.

2 Method

We are conducting empirical work using the C3Fire microworld [10]. Microworlds are simulated task environments that realistically capture aspects of a decision making problem. The origin of microworld research is Dörner & Brehmer's proposal that the investigation of intelligence should be conducted using simulated complex problem solving scenarios in the laboratory [11]. They proposed the use of microworlds, environments that (a) provide a task that can be made more complex, challenging, and realistic than traditional laboratory studies but that (b) generalize to interesting parts of real world problem solving while remaining (c) controllable and more easily analyzed than field study. Microworlds can be characterized along the dimensions of complexity, dynamics, polytely, and intransparency [12], [11], [13]. Complexity refers to the number of variables and causal relations in the task [11]. Dynamics refers to the degree of autonomous change of the controlled system, polytely to the degree of necessity of making trade-offs between multiple potentially conflicting goals, and intransparency to the fact that not everything in the system is directly observable. By enabling the experimenter to manipulate these dimensions, microworlds provide a much more tractable, reproducible, and flexibly designable research environment than a field study [14, 15]. Moreover, microworlds allow for the investigation of questions that cannot be addressed in either field studies or laboratory experiments [15], and at the same time they can address issues that can be seen in real world tasks [13]. Experimental participants take well designed microworlds seriously and become so engaged that their behaviour becomes completely natural and, accordingly, valuable to the researcher [15], [16]. These considerations make microworlds appropriate test beds for studying the propagation of constraints on plans made by decision makers working in a dynamic environment.

2.1 C3Fire

The C3Fire microworld has been used in previous research on network-based command and control [10], [17]. Figure 1 presents an overview of the C3Fire microworld.

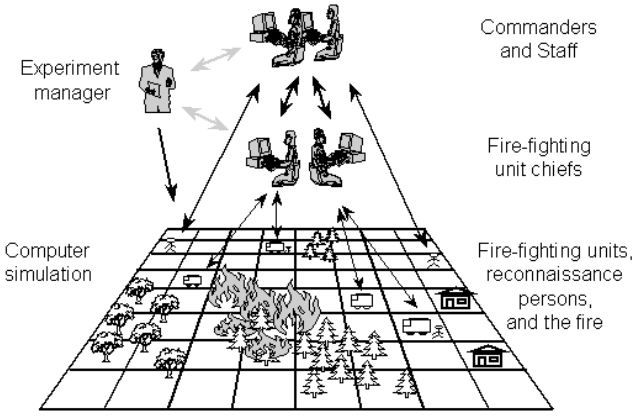


Figure 1. The C3Fire microworld

C3Fire [10] is a fire-fighting microworld in which a group of people collaboratively fights fires in an experimentally controlled configuration for command and control interaction, under observation of an experiment manager. In this microworld various entities and agents are mutually constraining.

The speed of burning and spreading of the fire can be determined by the experiment manager, and is typically made dependent on the vegetation of the terrain and presence of buildings, and the wind direction and speed. Thus, these properties of the environment constrain the development of the fire. The terrain and the developing fire are displayed on a Geographical Information System (GIS).

The fire is extinguished by placing fire-fighting units (fire-trucks) on top of the squares that represent fires on the GIS. Their movement is constrained by limits on their driving, deployment, and extinguishing speeds, set by the experimenter. The fire-fighting unit chiefs direct the units where to drive. The experimenter can limit the visual range of the units and constrain the information of the simulated world available to the unit chiefs to the information that is within visual range of the units under their direct control. Unit chiefs are constrained in their perception and action by the properties of the units that they control, by the visibility that the units give to them, and by the information and commands that they receive from commanders and staff. The commanders and staff can constrain and shape the action of fire-fighting unit

chiefs by communicating information about suitable positions for fire-trucks, locations of developing fire, and any message necessary.

In previous studies of network-based command and control using C3Fire, the fire trucks were the only class of unit under the fire chiefs' control (e.g., [17]). This restriction facilitated analysis of the effects of different configurations for command and control interaction on the decision making process. Our focus is less on the effects of organizational structure than on the interplay of constraints on the plans formulated by a distributed group of decision makers. To this end, we and others using C3Fire are extending it to include additional command functions and five new classes of units.

Two of the new classes of units are water and fuel trucks. Water trucks supply fire-fighting units with water, and obtain water from water stations (e.g. lakes). Fuel trucks supply both fire trucks and water trucks and obtain fuel from storage dumps located far from the fire. The addition of water and fuel trucks close constraints on the problem solving process by limiting the availability of fire trucks and by circumscribing the routes that fire trucks must take as they mobilize. In addition, each type of truck has characteristics that constrain its performance, i.e., mobilization velocity and tanking rates. The water and fuel dumps introduce additional constraints. These limited resources must be allocated across units and cannot be moved once established. They can, however, be consumed by the fire. The subgoal of protecting limited resources can lead some decision makers down a garden path.

The other new classes of units to be controlled by the collaborating decision makers are helicopter fire-fighting units, rescue teams, and civilians. The helicopters have essentially the same properties as the fire trucks but carry less water and move and deploy faster. Because the helicopters have a relatively high velocity, they can be used for reconnaissance in addition to fire-fighting. This dual purpose can introduce the need to open a closed constraint on their function. The rescue crews are able to rescue citizens from structures threatened by the approaching fire. Their vehicles require fuel that could otherwise be used to support fire trucks. The addition of civilians introduces a source of uncertainty into the system. A random number generator determines the willingness of the civilians to be rescued. Some want to stay in their cabins, hindering rescue efforts and raising unforeseen contingencies.

Interdependencies among decision makers arise whenever different classes of units are assigned to different participants in the simulated command and control center. For example if fire-trucks need water and fuel, water trucks and fuel trucks constrain the actions of fire-trucks. If different people have control over these different units, their actions constrain each other.

Even without these additional sources of constraint, the C3Fire microworld captures many of the characteristics of complex environments: It implements high degrees of connectivity, complexity, uncertainty, time pressure, and polytely. It provides a test bed for research on collaborative and distributed command and control and its interaction with the dynamics of a volatile, uncertain environment. The user interface of C3Fire provides access to a Geographical Information System, a diary, and an e-mail system. Together with the simulation of the fire fighting world itself, these interface systems are sources of operational, collaborative, and personal work information [10].

Previous research on distributed command and control has benefited from C3Fire's implementation of monitoring techniques in the form of a knowledge management system that analyzes the information generated by the spreading fire, the chiefs and commanders, and by human-machine and human-human interaction. The project discussed in this paper aims to enhance these monitoring tools and aims to use the information extracted from the interactions to support our analysis of constraint propagation and the compatibility of plans.

The classes of constraints on the compatibility of plans that we plan to investigate using C3Fire are (a) the constraints that the actions (current and proposed) of fire-fighting unit chiefs and commanders impose on each other's actions, (b) the constraints imposed by the availability in time and space of vehicles or other useful tools and objects, (c) the constraints that the physical world (the fire) and social world (the dispositions of civilians) impose on the actions of fire-fighting units and commanders, (d) the constraints of uncertain information and distributed knowledge, and (e) the constraints that the control of single fire-fighting units by several unit chiefs and commanders impose on actions of decision makers and vehicles. The opening and closing of constraints and the interdependence among participants are likely to manifest themselves in C3Fire through (a) the choice and flexibility or rigidity of communication and command and control architectures, (b) the deployment of different types of vehicles, (c) the terrain and time that define the initial state.

1.1 Constraint propagation and planning

Planning is the task of charting a course through the problem space. The two major sources of constraint on planning are the availability of time and information [18]. When we plan, we sometimes allow closed constraints to proliferate and, in the process, we narrow the problem

space and make our plans rigid and strict. At other times, when it suffices to have only a rough idea of what to do, we allow some constraints to remain open [8]. Accordingly, plans can take many forms, from the very explicit in representations of tables with time, location, and action, to the most implicit in vague ideas, intentions, and descriptions of actions.

A plan is a system of constraints on action. To assess how plans constrain each other and collaborative decision making, data have to be captured that reveal plans as they unfold in an uncertain world. These data have to be analyzed with a certain method to identify constraints set by these plans, and how they act on other plans, either enabling or restricting them. The classes of constraints and their data instances are discussed below, together with methods under investigation to detect them.

Every event in the simulation generates time-stamped data. Unit movements, actions of decision makers, and communications between decision makers form a vast amount of data that is generated and stored. The current project aims to analyze these data and extract plans and constraints that decision makers implicitly or explicitly use in solving their problems. The data available for such analysis are to result in certain kinds of knowledge that inform the detection of plans:

(a) Knowledge about the decision makers' capabilities. Each decision maker can have different capabilities than the others, corresponding to the type and amount of vehicles they control, and their possibilities for communication. Decision makers may have the same capabilities or the same units available to them, corresponding to the network-based command and control architecture where the locus of control can change.

(b) Knowledge about the actions of the decision makers. The intentions of controlling vehicles are known if and when the decision makers place their markers that indicate where the vehicles are supposed to move.

(c) Knowledge about the world: (1) The fire, where it is, since when, if it is burning, burnt down, or closed out. (2) The properties of the terrain and environment of the world. For example, to a certain extent one knows how fast the fire spreads in certain vegetation, how fast and where units can travel. (3) The anticipated behavior of the fire. Assumptions can be made based on more or less uncertain information on how the fire will develop in the future. For example, the wind direction and speed form a basis for estimating where the fire will develop and at what speed. (4) The vehicles moving around to fight the fire and their parameters, locations, destinations, and activities.

(d) Knowledge about the knowledge available to the decision makers. Each decision maker has a unique view of the world, in which certain information is presented about the world, and about the other actors in the world.

(e) Knowledge about the orders, instructions, and information that is communicated between commanders, staff, and unit chiefs. Currently the commanders' messages are in natural language, which could be parsed, but more suitably the ways of communication can be automated so that communication is efficient (one does not need to type) and orders and instructions can be analyzed.

(f) Knowledge about the strategies may become available when repeated experimentation is performed and more or less clear strategies can be observed.

1.2 Intelligent interface design

The knowledge described above is to be used to identify constraints and plans. Our aim is to support decisions through an intelligent interface that explicitly represents constraints, which will highlight opportunities and impossibilities for action for the decision makers. It is important to note that the focus is on supporting intelligent decisions that are to be made by the human decision makers. They can however be aided by intelligent systems in their decision making, and adaptive interface technology and multimodal interaction (see also for example [19], [20]) in combination with constraint propagation is under investigation as the foundation for effective and efficient decision support. Representation design and ecological interface design goals are taken as a basis here, summarized by (a) managing information to avoid overload, (b) facilitation of the discovery of constraints, incompatibilities, and, thereby, patterns of meaning, (c) clearly representing constraints in a way that makes the possibilities for action and resolution evident, (d) highlighting dynamic change, whether of events, organizational structure, or constraint, (e) functioning as an external memory to the decision maker, (f) directing attention in an appropriate way, and (g) creating a system that is robust in the face of unexpected situations [21], [22], [23].

References

[1] H. Friman, "Decision making in a ROLF staff environment," in *ROLF 2010 - The Way Ahead and The First Step*, *Försvarshögskolans Acta C6*, C. Sundin and H. Friman, Eds. Stockholm: The Swedish National Defence College, 2000, pp. 67-74.

[2] B. Brehmer and C. Sundin, "Command and control in network-centric warfare," in *ROLF 2010 - The Way Ahead and The First Step*, *Försvarshögskolans*

Acta C6, C. Sundin and H. Friman, Eds. Stockholm: The Swedish National Defence College, 2000, pp. 45-54.

[3] B. Brehmer, "Distributed decision making: some notes on the literature," in *Distributed Decision Making: cognitive models for cooperative work*, J. Rasmussen, B. Brehmer, and J. Leplat, Eds. Chichester: John Wiley & Sons Ltd, 1991, pp. 3-14.

[4] B. Brehmer, "Dynamic decision making: Human control of complex systems," *Acta Psychologica*, vol. 81, pp. 211-241, 1992.

[5] K. E. Weick and K. H. Roberts, "Collective mind in organizations: Heedful interrelating on flight decks," *Administrative Quarterly*, vol. 38, pp. 357-381, 1993.

[6] Z. Shapira, Ed., *Organizational Decision Making*. New York: Cambridge University Press, 1997.

[7] A. Newell and H. A. Simon, *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972.

[8] W. Reitman, R., "Heuristic decision procedures, open constraints, and the structure of ill-defined problems," in *Human Judgments and Optimality*, M. W. Shelly II and G. L. Bryan, Eds. New York: John Wiley & Sons, Inc., 1964, pp. 282-315.

[9] E. Hutchins, "Organizing work by adaptation," *Organization Science*, vol. 2, pp. 14-39, 1991.

[10] R. Granlund, "Monitoring experiences from command and control research with the C3Fire microworld," *Cognition, Technology & Work*, vol. 5, pp. 183-190, 2003.

[11] J. Funke, "Dynamic systems as tools for analysing human judgement," *Thinking and Reasoning*, vol. 7, pp. 69-89, 2001.

[12] J. Funke, "Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results," in *The Cognitive Psychology of Knowledge*, G. Strube and K. F. Wender, Eds. Amsterdam: Elsevier, 1993.

[13] G. Rigas, E. Carling, and B. Brehmer, "Reliability and validity of performance measures in microworlds," *Intelligence*, vol. 30, pp. 463-480, 2002.

[14] B. Brehmer and D. Dörner, "Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study," *Computers in Human Behavior*, vol. 9, pp. 171-184, 1993.

[15] W. D. Gray, "Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research," *Cognitive Science Quarterly*, vol. 2, pp. 205-227, 2002.

[16] D. Dörner and H. Schaub, "Spiel und wirklichkeit: Über die verwendung und den nutzen computersimulierter planspiele," *Kölner Zeitschrift für Wirtschaft und Pädagogik*, vol. 12, pp. 55-78, 1992.

[17] B. Johansson, M. Persson, R. Granlund, and P. Mattsson, "C3Fire in command and control research," *Cognition, Technology & Work*, vol. 5, pp. 191-196, 2003.

[18] G. A. Miller, E. Galanter, and K. H. Pribram, *Plans and the Structure of Behavior*. New York: Holt, Rinehart and Winston, Inc., 1960.

[19] K. Höök, "Steps to take before intelligent user interfaces become real," *Interacting with Computers*, vol. 12, pp. 409-426, 2000.

[20] M. F. McTear, "Intelligent interface technology: from theory to reality?," *Interacting with Computers*, vol. 12, pp. 323-336, 2000.

[21] D. D. Woods, "Toward a theoretical base for representation design in the computer medium: Ecological perception and aiding human cognition," in *Global Perspectives on the Ecology of Human-Machine Systems*, *Resources for Ecological Psychology*, J. Flach, P. Hancock, J. Caird, and K. Vicente, Eds. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 1995, pp. 157-188.

[22] E. M. Roth, J. T. Malin, and D. L. Schreckenghost, "Paradigms for intelligent interface design," in *Handbook of Human-Computer Interaction*, M. Helander, T. K. Landauer, and P. Prabhu, Eds., 2nd ed. Amsterdam: Elsevier Science, 1997, pp. 1177-1201.

[23] K. J. Vicente and J. Rasmussen, "Ecological interface design: Theoretical foundations," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, pp. 589-606, 1992.

HUD USE IMPROVES LANDING PERFORMANCE IN A FULL-MOTION SIMULATOR

Örjan Goteman* #, Kip Smith*, and Sidney Dekker*
*Linköping Institute of Technology, Sweden
Scandinavian Airlines, Sweden

The operational community has assumed that using a head-up display (HUD) rather than conventional head-down displays will increase safety margins in approach and landing operations. This paper discusses a pair of experiments that test that hypothesis. Using a Boeing 737-700 full-motion simulator and commercial airline pilots as subjects, we explored the effects of (1) HUD use, (2) ambient visibility, and (3) length of approach lighting on the size and location of the touchdown footprint. HUD use reduced the width of the touchdown footprint in all tested visibility and lighting conditions including visibility below the minimum allowed. HUD use had no effect on the length of the touchdown footprint. These are empirical data that could inform the writing of regulations regarding HUD-supported commercial aviation operations.

INTRODUCTION

This paper reviews two experiments designed to assess the impact of commercial airline pilots' use of a *head-up display* (HUD) on approach and landing performance under reduced runway visual range (RVR) conditions. We begin by reviewing previous research on HUD use and by describing RVR approach and landing operations. The review motivates a pair of hypotheses regarding the impact of HUD use on the size of the landing footprint. The experiments and their findings are discussed in turn.

HUD use in approach and landing operations

The HUD projects aircraft flight-path and energy symbols onto a transparent screen. Pilots can concurrently view the HUD symbology and the surrounding world. *Conformal* HUD displays preserve a 1:1 relationship between the visual angles of displayed symbols and their real world referents. Hoerner (1979) reported a threefold reduction in landing accidents on aircraft carriers when pilots used a conformal HUD with a flight-path symbol. Comparisons between HUD and head-down displays in manual flight found that HUD use improves track, speed, and altitude maintenance (Lauber et al., 1982; Martin-Emerson & Wickens, 1997).

The civil operational community has assumed that HUD use mitigates two well-documented problems associated with approach and landing: (1) visual approaches to runways without radio navigation aids or with unreliable navigation aids, and (2) the transition from instrument to external visual cues for landing in low visibility (e.g., Newman, 1995). To date, these assumptions appear to have gone untested.

Approach operations in reduced visibility

The majority of *approaches* (the segment of flight immediately before touchdown) performed in US and European civil air transport operation are conducted as Category I instrument landing system (ILS) approaches. Two critical factors when making the decision of whether or not to land are the *decision height* and the *runway visual range* (RVR). Decision height refers to the aircraft's vertical distance above the runway threshold at the point where the pilot must make that decision. The RVR is a measure of horizontal visibility defined by the number of visible approach and runway lights in the ambient atmospheric conditions.

To begin a Category I ILS approach, the RVR at the decision height must not be less than 550 m (International Civil Aviation Organization [ICAO], 2001). Operationally, the pilot needs to be able to see from decision height a visual segment of the ground that contains enough of the runway approach lights to judge the aircraft's lateral position, cross track velocity, and position in roll (Joint Aviation Authorities [JAA], 2003). This condition is sketched in Figure 1. As the aircraft comes closer to the runway threshold, the pilot sees and aims at the designated touchdown point 305 m down the runway. If the RVR is too low at the decision height, the pilot will not be able to see some or any of the approach lights and must make a go-around. Although the resulting *missed approach* is part of normal operations (and formal procedures), it adds an undesired additional risk (ICAO, 2003).

The issue of RVR and HUD use

Commercial airlines are introducing HUDs to the flight deck at an increasing rate, making it beneficial to know the effect of HUD use on touchdown performance. Generalizing from military research (Hoerner, 1979), HUD use with conformal symbology should improve tracking performance during approach and landing in restricted visibility. However, there appear to be no empirical data to support such a claim. Data do exist for certification trials for HUDs currently used in civil aviation but those trials used the HUDs in a low visibility mode, in which there was no need for external visual cues (JAA, 2003). Such data do not export well to touchdown performance using HUDs in normal and reduced visibility flight operations where obstacle clearance below the decision height is maintained by visual maneuvering. As a result, regulators are being pressed to generate rules that are insufficiently grounded in research.

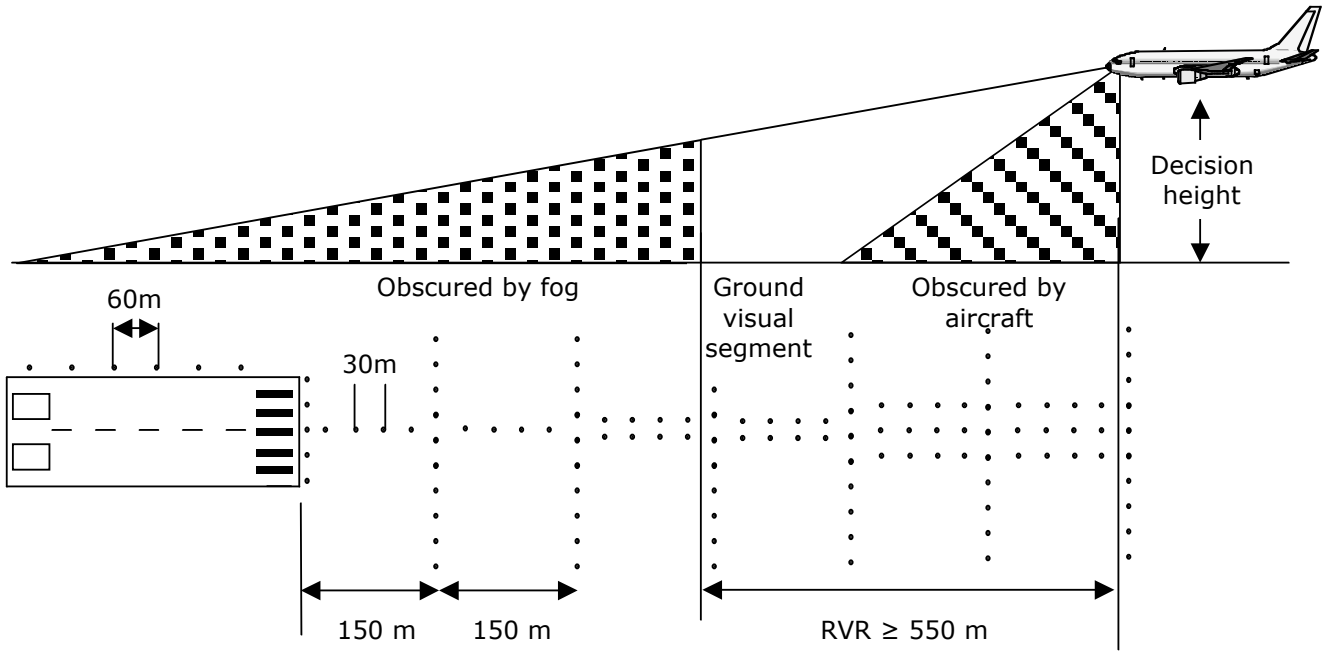


Figure 1. The visual segment seen from the decision height given the runway visual range (RVR).

In light of the increasing use of HUDs in commercial aviation and the paucity of the existing knowledge base, we studied whether the inferred improvement in tracking performance when using a HUD will be reflected in a smaller touchdown footprint on the runway. We predicted that the HUD would reduce the size of the footprint both longitudinally and laterally.

EXPERIMENT 1

We had initially considered conducting a series of experiments that would fully cross manipulations of RVR and the length of approach lighting to assess the interaction of these factors on HUD use during approach and landing operations. It immediately became obvious that not all combinations needed to be studied. For example, experienced commercial pilots are adept at making approaches and landings at facilities with a full system of approach lights (900 m) when visibility is above the RVR minimum of 550 m. We also found that the majority of pilots would opt for a missed approach rather than attempt to land at facilities with less than a full system of approach lights when visibility was below the RVR minimum. Accordingly, we conducted two experiments. In the first, we assessed HUD use in visibility below the RVR minimum at simulated facilities with a full system of approach lights. In the second, we assessed HUD use in visibility above the RVR minimum at simulated facilities with a system of approach lights of intermediate length (420 or 600 m).

Method

Participants. Forty-three pilots from a major European airline volunteered to participate. All were qualified to fly the B737-700 aircraft and had completed their HUD training for the operator. The training sessions consisted of one day of theory and two simulator sessions of four hours duration each. Experience on the B737-700 varied from 50 to more than 1000 hours. There is every reason to believe that these participants are representative of the population of commercial pilots to whom regulators need the data to generalize.

Apparatus. We used a CAE B737-700 training simulator with aircraft aerodynamics and visual angles valid for B737-700 to collect data. This six-axis full-motion simulator is approved for low-visibility operations down to an RVR of 200 m. The pilot had a field of vision of 180°/40° collimated to a focal distance greater than 10 m.

The HUD installed in the simulator was a Rockwell-Collins Flight Dynamics HGS-4000® (Head-up Guidance System), certified for low visibility operations down to and including RVR of 200 m. The HUD symbology used in the study is shown in Figure 2. The symbology and functionality met the production-line standard specification for the mode customarily used (IMC, instrumental meteorological condition mode) when making Category I instrument landing system (ILS) and non-precision approaches. The HUD provided conformal display of flight path and flight-path guidance. Flight-path guidance is not available in the conventional head-down instrumentation of the B737-NG with a primary flight display. The HUD provided the flight-path guidance using a ring inside the flight-path symbol.

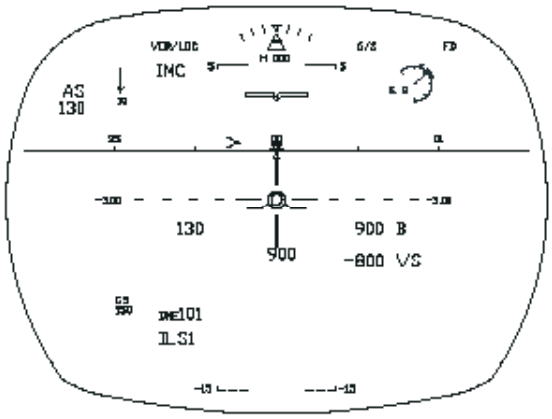


Figure 2. HUD symbology in the IMC mode used for Category 1 ILS and non-precision approaches.

Using a HUD mode with no flare guidance improved the external validity (in the form of operational usability) of the study. In standard operational settings, the vast majority of the ILS systems are approved for only Category I operations and do not have the quality and integrity to support HUD modes incorporating landing-flare guidance.

Simulated ground facilities. The ground facilities simulated in the study conformed to the International Civil Aviation Organization (ICAO, 1996) standard for ILS radio navigation aids. Such a system contains tracking, approach, and runway lights and transmits a beam for both vertical and lateral reference. Approaches in Experiment 1 were flown to a simulated runway with an system of approach lights 900 m in length. This length falls within the *full system* category of European aviation regulation (JAA, 2003).

Design. The experiment used a complete block design with repeated measures on HUD use (HUD, no-HUD). The order of HUD use was counterbalanced across participants to control for potential order and carryover effects. Two blocks of visibility, the standard minimum RVR (550 m) and lower-than-standard RVR (400 to 500 m), were a between-subjects variable. Each participant completed two successful approaches to a runway with a full system (900 m) of approach lights.

Procedure. Two approaches were manually flown using standard operating procedures. The scenarios started as a 6 nautical mile final to the runway in lower-than-standard or standard RVR. In the with-HUD condition, the pilots flew the aircraft by keeping the flight-path symbol centered on the guidance cue. At 50 m above the runway threshold, the guidance cue was automatically removed and the pilots performed the landing flare using external visual cues in conjunction with the HUD flight-path symbol. The automatic removal of the guidance cue and the HGS-4000® IMC mode were deliberately chosen to insure that the pilots could not attend solely to the HUD symbology in the with-HUD trials. Approach and landing plots were printed using the aircraft's center of gravity as reference to determine the size of the touchdown footprint.

Results

A series of full-model and reduced-model linear regressions were calculated to create separate ANOVA tables for the lateral (cross-runway) and longitudinal (down-runway) components of the touchdown footprint. The regression approach is robust to differences in sample sizes across blocks (levels of RVR) and allows for assessment of order effects and their interactions with the main factor of interest, HUD use.

Lateral touchdown performance was measured as the absolute lateral deviation from the runway centerline at touchdown. The data for the main effect of HUD use is shown in Figure 3a. There was a strong effect for HUD use, $F(1, 78) = 8.991$, $p < .004$, $\eta^2 = .12$. Landings were closer to the centerline when pilots used the HUD. The main effects for RVR and order of HUD condition were not significant. There was, however, a marginally significant interaction between RVR and the order of HUD use, $F(1, 78) = 3.235$, $p < .076$, $\eta^2 = .11$. As shown in Figure 3b, both the group that used the HUD first (HUD 1st) and the group that used the HUD after using the head-down display (no HUD 1st) landed nearer the centerline when they used the HUD. The improvement is greater for the group that used the HUD in the second approach.

Longitudinal touchdown performance was measured as the distance from the runway threshold at touchdown. Contrary to expectations and in contrast with the results for lateral performance, the effect of HUD use was not significant. The effect of RVR, shown in Figure 4a, was significant, $F(1, 78) = 6.504$, $p < .013$, $\eta^2 = .07$, as was the interaction of RVR and group, $F(1, 78) = 5.750$, $p < .019$, $\eta^2 = .13$, Figure 4b. Landings were considerably shorter in low RVR for the group that used the HUD first.

EXPERIMENT 2

Method

The method, procedure, and design used in the second experiment were identical to those used in the first experiment with the few exceptions discussed here. Large differences in sample size across the experimental groups and the blocks of RVR make it unwise to attempt to collapse and analyze the data as a single experiment.

Participants. 28 pilots from the same major European airline volunteered to participate. None of the volunteers had participated in Experiment 1. All were qualified to fly the B737-700 aircraft and had completed their HUD training for the operator. Experience on the B737-700 varied from more than 50 to more than 1000 hours.

Simulated ground facilities. Approaches in Experiment 2 were flown to simulations of the runways at Bruxelles (ILS 25R BRU) with 600 m of approach lights or at Tromsö (ILS 01 TOS) with 420 m of approach lights. These lengths fall within the *intermediate* category of approach lights as defined by European aviation regulation (JAA, 2003).

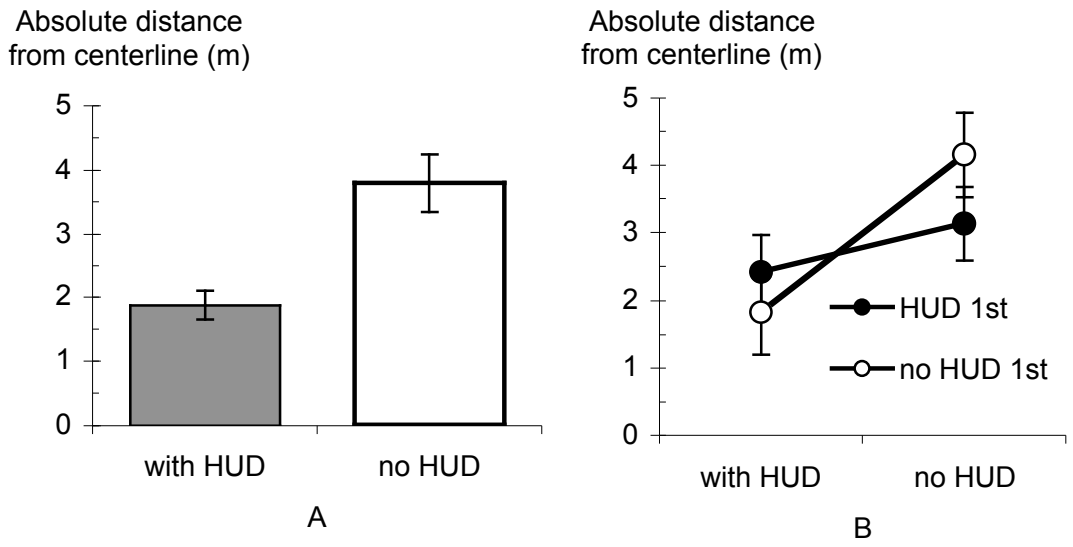


Figure 3. Graphs showing (A) the main effect for HUD use and (B) the interaction of HUD use and the order of its use on the lateral size of the touchdown footprint at simulated facilities with a full system (900 m) of approach lights, Experiment 1.

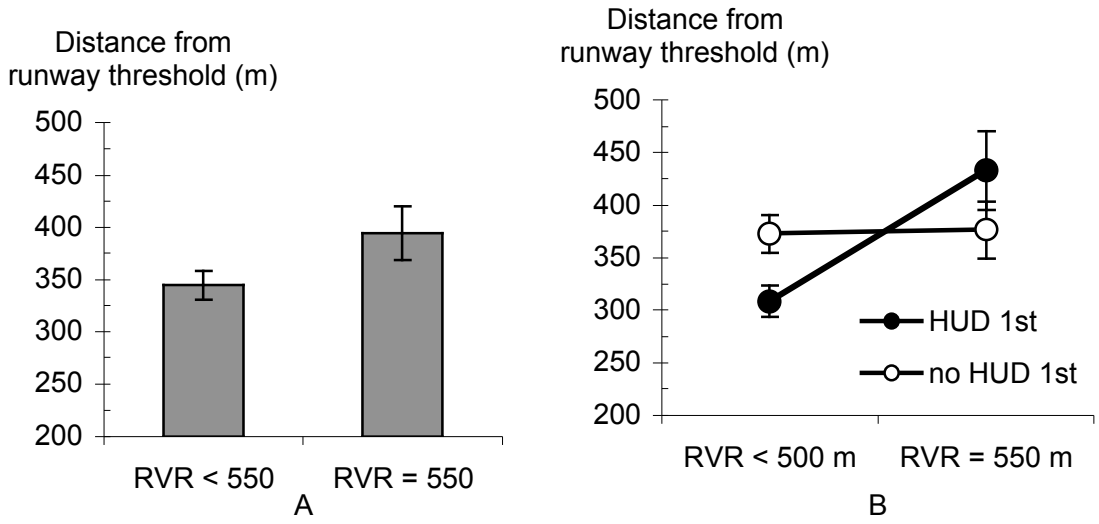


Figure 4. Graphs showing (A) the main effect for RVR and (B) the interaction of RVR and trial order of HUD use on the longitudinal size of the touchdown footprint at simulated facilities with a full system (900 m) of approach lights, Experiment 1.

Design. Two blocks of visibility, the standard minimum RVR (550 m) and greater-than-standard RVR (600 to 700 m), were a between-subjects variable in a complete-block repeated-measures design. Each participant completed two successful approaches to a runway with a system of approach lights either 420 or 600 m in length.

Results

No differences were found across the two intermediate lengths of runway lights. The analyses presented here collapse the data across this variable. As in Experiment 1, there was a strong effect for HUD use, $F(1, 48) = 7.380$, $p < .009$, $\eta^2 = .09$. The data for the main effect of HUD use is shown in Figure 5. Once again, landings were closer to the centerline when pilots used the HUD. The main effects for RVR and order of HUD use and all interactions on lateral touchdown performance were not significant. None of the factors in the ANOVA for longitudinal touchdown performance approached significance.

DISCUSSION

HUD use resulted in a significantly smaller lateral touchdown footprint for all RVR conditions. In contrast, it had no effect on the size of the longitudinal footprint. A plausible explanation for this differential result lies in the

interaction of pilot adaptation and the nature of the additional information afforded by the HUD.

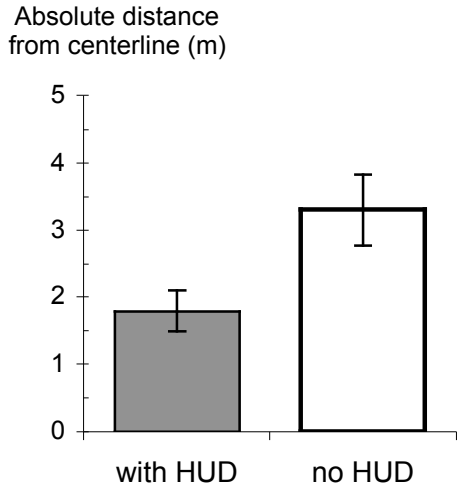


Figure 5. Graph showing the main effect of HUD use on the lateral size of the touchdown footprint at simulated facilities with an intermediate system (420 m or 600 m) of approach lights, Experiment 2.

Experience engenders automaticity. For commercial pilots, hours of practice and numerous repetitions make touchdowns at standard facilities in good visibility and weather relatively routine. Nevertheless, the ubiquitous and ever-varying direction and velocity of crosswinds are likely to preclude the development of true automaticity at touchdown. Crosswinds introduce an element of uncertainty regarding *drift* (the lateral location of the aircraft relative to the centerline.) This exogenous source of uncertainty is likely to have trained pilots (and regulators) to tolerate a certain amount of variance in the lateral touchdown footprint.

The HUD largely eliminates uncertainty about drift. The addition of a conformal flight-path vector provides instantaneous feedback about aircraft drift and actual flight path. It compensates for the pilots' learned tolerance for lateral drift in the landing footprint.

In contrast, it appears that the HUD does not influence the size of the longitudinal footprint. Variance in the longitudinal footprint is largely an effect of how pilots handle the aircraft's energy in the final seconds before landing (its *sinkrate*). The HUD provides information that enables precise control of the flight path during approach but does not provide information about sinkrate. (To facilitate control of sinkrate, large commercial air transport aircraft are equipped with a radio altimeter that counts down from 50 feet to 0 feet (runway contact) in 10 foot increments.) The HUD does not influence the size of the longitudinal footprint because the observed variance in performance is largely determined by variables other than precise control of the flight path.

However, our data suggest that HUD use and low RVR conditions are likely to interact and influence the point of touchdown. In Experiment 1, pilots consistently landed short of the designated touchdown point when using the HUD. We attribute this observation to a change in the focus of attention. During approach in low RVR, pilots follow their instruments; they focus their attention on the HUD symbology. It is likely that pilots focusing on HUD symbology assimilate outside cues later than pilots who have no HUD symbology. A later landing flare would lead to the shorter touchdown point that we observed in our experiment.

These results are the first of their kind that regulators can use when deciding upon the visibility required to commence an approach using a HUD-equipped aircraft. We believe that empirical results like these should be a critical ingredient in safety regulators' rule-writing for (the application of) new technology in commercial aircraft operations. The need for informed regulation of HUD use is already pressing. That need will only grow as HUD use grows. We encourage regulators to take these findings into account, to sponsor replication of this research paradigm, and to expand it to, e.g., Category II conditions.

ACKNOWLEDGEMENTS

The authors wish to thank Staffan Carlsson and Zacharias Soudah, both from Scandinavian Airlines for collecting and systemizing the data from the trials. Jan-Olof Ehk, Scandinavian Airlines, provided valuable support by giving statistical advice. Rockwell Collins Flight Dynamics Company, provided simulator time to conduct the experiment. Finally the authors wish to thank the pilots who volunteered their time and skill to this experiment.

REFERENCES

Hoerner, F. C. (1979) HUD military aspects. *Proceedings of the Flight Operations Symposium, Vancouver* **2**, 53-58, 117-142. Redmond, WA: Sundstrand Data Control.

ICAO. (1996). *Annex 10, Aeronautical Telecommunication, Volume I – Radio Navigation, 5th edition*. Montreal, Quebec: International Civil Aviation Org.

ICAO. (2001). *Annex 6, Operation of Aircraft, 8th edition*. Montreal, Quebec: International Civil Aviation Org.

ICAO. (2003). *OPS – Aircraft Operation (Doc 8168) Volume II, 4th edition*. Montreal, Quebec: International Civil Aviation Org.

Joint Aviation Authorities. (2003). *All weather operations*. Hoofddorp, The Netherlands: European Civil Aviation Conference.

Lauber, J. K., Bray, R. S., Harrison, R. L., Hemingway, J. C., & Scott, B. C. (1982). An operational evaluation of head-up displays for civil transport aircraft. *NASA Tech. Paper 1815*. Moffett Field, CA: NASA Ames Research Center.

Martin-Emerson, R. & Wickens, C. D. (1997). Superimposition, symbology, visual attention and the head-up display. *Human Factors* **39**(4), 581-601.

Newman, R. L. (1995). *Head-up displays: Designing the way ahead*. Aldershot: Ashgate.

Initial Experiments On Leader Presence And Communication Mode On Combat Performance

Kip Smith
Linköping Institute of Technology
Division of Industrial Ergonomics
581 83 Linköping Sweden

ABSTRACT

The US Army and NATO forces are in the process of shifting from the traditional in-situ mode of command and control between soldiers and their leaders to a distributed mode of command and control. As part of this shift, a fire-unit's leader may no longer be part of the unit on the battlefield. Rather, the leader may sit at a relatively remote location and use a variety of electronic media to communicate with the team. In the experiments discussed here, we are starting to address the impacts of remote command and control and communication mode in a series of ecologically realistic simulations of a battlefield environment. We have found that participants follow orders more quickly in the leader-present condition. This result suggests that some kind of intervention will be required if soldier performance is to be as efficient in remote command and control as it is in the more traditional, leader-present, mode of control.

Keywords: Remote command and control, Leader presence, Mode of communication, Combat

INITIAL EXPERIMENTS ON LEADER PRESENCE AND COMMUNICATION MODE ON COMBAT PERFORMANCE

Kip Smith
Linköping Institute of Technology

ABSTRACT

The US Army and NATO forces are in the process of shifting from the traditional in-situ mode of command and control between soldiers and their leaders to a distributed mode of command and control. As part of this shift, a fire-unit's leader may no longer be part of the unit on the battlefield. Rather, the leader may sit at a relatively remote location and use a variety of electronic media to communicate with the team. In the experiments discussed here, we are starting to address the impacts of remote command and control and communication mode in a series of ecologically realistic simulations of a battlefield environment. We have found that participants follow orders more quickly in the leader-present condition. This result suggests that some kind of intervention will be required if soldier performance is to be as efficient in remote command and control as it is in the more traditional, leader-present, mode of control.

Keywords: Remote command and control, Leader presence, Mode of communication, Combat

INTRODUCTION

The practice of having soldiers on the battlefield receive orders from afar through electronic means of communication is known as remote command and control. A reliance on remote command and control is one of the cornerstones of the US Army's plan for modernizing the dismounted infantry. The soldiers who will operate under this plan are (currently) known as the Future Force. With the advent of the Future Force concept, soldiers may no longer take their battle commands from a leader standing within visual range. Instead, the only connection with their commanding officers may be their radios and other portable information devices.

Previous research has shown that varying the physical proximity of an authority figure affects a person's compliance with a command. In the classic study by Milgram (1974), a research participant was far more likely to administer electric shocks to another person at the researcher's command if the researcher was present. If the researcher gave an order to punish an individual from a separate room via telephone, the participant was three times less likely to comply with the command than if the researcher were in the room giving the command. Accordingly, it is reasonable to hypothesize that a change from leaders who are present on the battlefield to leaders who give orders from a distance is likely to have an adverse impact on soldier performance. The study discussed here investigates the effect of leader presence at two levels (present vs. remote) on soldiers' response to commands to move and to shoot. We anticipate that remote command and control will degrade a leader's ability to exercise authority. We expect this degradation in perceived authority will be reflected in slower reaction times and higher levels of psychophysiologic stress when commands are given remotely and when given over a radio than when they are given face-to-face. If this is found to be the case, it will be necessary to design interfaces and training regimes to insure that this degradation of authority can be mitigated.

METHOD

In the set of three experiments presented here, we have modified the Milgram task to make it palatable to institutional review boards and to give it sufficient ecological validity to generalize to a military setting. The technology that enables this simultaneous ethical sanitization and realism is called Paintball.

The first two experiments focused exclusively on behavioral measures and on the effect of leader presence (Pangburn, Freund, Pangburn, & Smith, 2003). Pangburn et al. document the utility of the paintball assault lane as an experimental platform for studying performance under live fire. The third study is in progress. It builds upon the first two to assess the potential for an interaction between leader presence and communication mode. It augments behavioral measures with analyses of two psychophysiologic indicators of stress - heart-rate and heart rate variability.

This section describes elements of the experimental method shared by all three experiments. Each experiment and its results are discussed separately.

Design, Measures, and Task

The simulated combat environment used in this study is a paintball assault lane, Figure 1. Participants advanced through the lane one at a time. The lane consisted of eight protective stations behind which the participant could hide. At the end of the lane was a fortified position where a sniper was positioned. The sniper's task was to shoot the participant moving up the lane. The participant had two tasks. The first was to move from station to station up the lane in response to the command to move. The second was to shoot targets in response to the command to shoot.

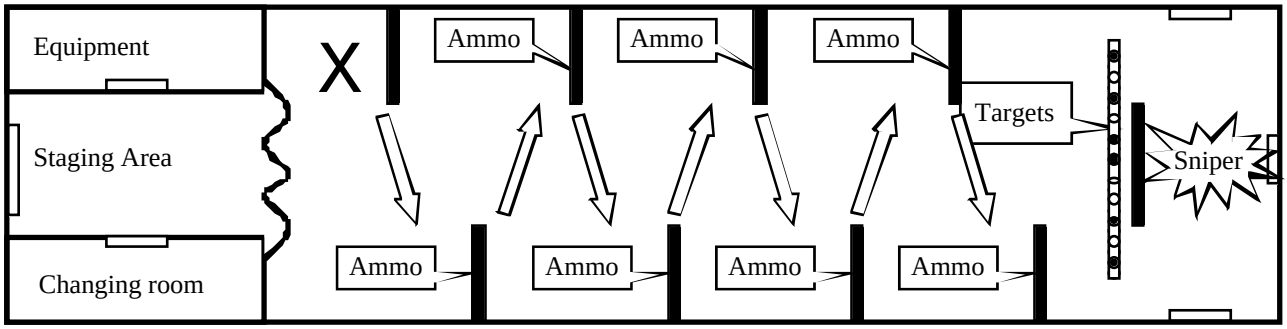


Figure 1 - The layout of the paintball assault lanes.

In all three experiments we manipulated leader presence at two levels (present and remote) in a within-subjects design. In the leader-present condition, the leader was one station behind the participant and communicated by yelling. In the leader-remote condition the only contact between the leader and the participant was by two-way radio.

We used a repeated-measures design with the order of conditions counterbalanced across participants. This design provides the statistical power needed to assess the effect of leader presence and mode of communication on the time it takes participants to respond to commands to move and to shoot.

We measured the participant's response time to the leader's commands to move and to shoot. We predict slower response times in the leader-remote condition but have no a priori hypotheses regarding the effect of communication mode. Statistical analysis used ANOVA to test for sequence effects and within-subjects t tests of the mean differences in response times.

Materials

Participants and the sniper were given one paintball marker (gun), fatigues (overalls), a set of elbow and kneepads, and a paintball face shield. In the leader-remote and present-radio conditions, participants were also given a two-way radio. In the first two experiments, response times to commands to move and shoot were recorded by an observer using a stopwatch. Procedure

Upon arriving at our lane the participants met the leader for the first time. The leader was an army officer wearing a standard battle-dress uniform. The leader briefed participants using the official military Operations Order format and addressed them by their last names. After signing informed consent and liability release forms, participants were told to assemble in a staging area where they could hear the activity in the assault lane while they waited their turns. While waiting,

participants were instructed on the safety and use of the paintball markers and read a briefing. All of this was purposefully done to immerse the participant in the experiment and to heighten the sense of realism and their anxiety.

Participants were sent down the lane individually. Whenever the participants took aim at the targets or moved between stations, they exposed themselves to the sniper's fire. Participants were instructed to attempt to shoot enemy targets without hitting friendly targets. No measures were made of firing accuracy, however, because our hypothesis concerns the participant's reaction time to commands given by the leader. The shooting task was created only to give focus to the participant's activity and to give the experiment the feel of a combat environment. (Post-experimental conversations suggest that shooting accuracy was strongly correlated with hunting experience.)

A small container with five paintballs was placed at each of the eight stations. The 40 paintballs in the eight containers were the participant's only ammunition. The participant started at one end of the lane, shown by the X in Figure 1. At this station and all subsequent stations, the leader gave the participant the command "Fire" when it appeared to be safe to do so. The time elapsed from the issue of the command to the first shot fired is the first dependent measure.

When the participants ran out of paintballs, they reported "Out of ammo" to the leader, who then gave the command "Move" when it appeared to be safe to do so. The participants had to move across the lane to the next station and immediately pick up its container of five paintballs. The time elapsed between the issue of the command to move and the time the participant's hand first touched the new supply of ammunition was the second dependent measure. When the participants finished loading, they reported "Loaded" to the leader who then started the cycle over again with the command "Fire."

The study was intended to generate some anxiety so that the measures would more readily generalize to the battlefield. The major sources of stress were the fear of being shot and actually being hit by paintballs. The pain associated with being struck by a paintball is slight but real. Protective gear minimized the risk of injury.

EXPERIMENT 1

Location and Participants

For the first experiment, the US Army provided access to the 25 ft x 200 ft building at Range 52, Fort Riley, Kansas, home of the US Army 1st Infantry and an active training center for artillery. We set up our paintball assault lane in this building.

Twenty volunteers from Kansas State University (18 men, 2 women; median age 19, range 18 to 28) participated in the first experiment. Attendance was limited because a one-way trip to Fort Riley took 45 minutes and required passing through a security gate and a variety of active firing range complexes. All told, the experiment took at least four hours of the participants' time.

Results

Figure 2a is a graph of response times to the command to move. The open symbols show the means and standard errors of response times for the group of participants who first ran the lane in the remote-leader condition. This group responded more quickly in the second trial when the leader was present in the lane. The closed circles, for the group who first ran in the leader-present condition, show that participants responded more quickly in the first trial, again when the leader was present in the lane. A two-factor ANOVA was conducted to assess sequence and group effects. Neither group (remote-first, present-first), $F(1,36) = .221$, $\underline{p} > .64$, nor sequence (first trial, second trial), $F(1,36) = .064$, $\underline{p} > .80$, were significant. This result allows us to merge data across groups and to conduct a within-subjects t-test for the effect of leader presence. The test, $t(19) = 2.958$, $\underline{p} < .004$, indicates that, as expected, leader presence made a significant difference in the participants' response times to commands to move. Cohen's d as adjusted for the lower variability inherent in a

repeated-measures design at an alpha of .05 is approximately .94, indicating ample statistical power with 20 participants.

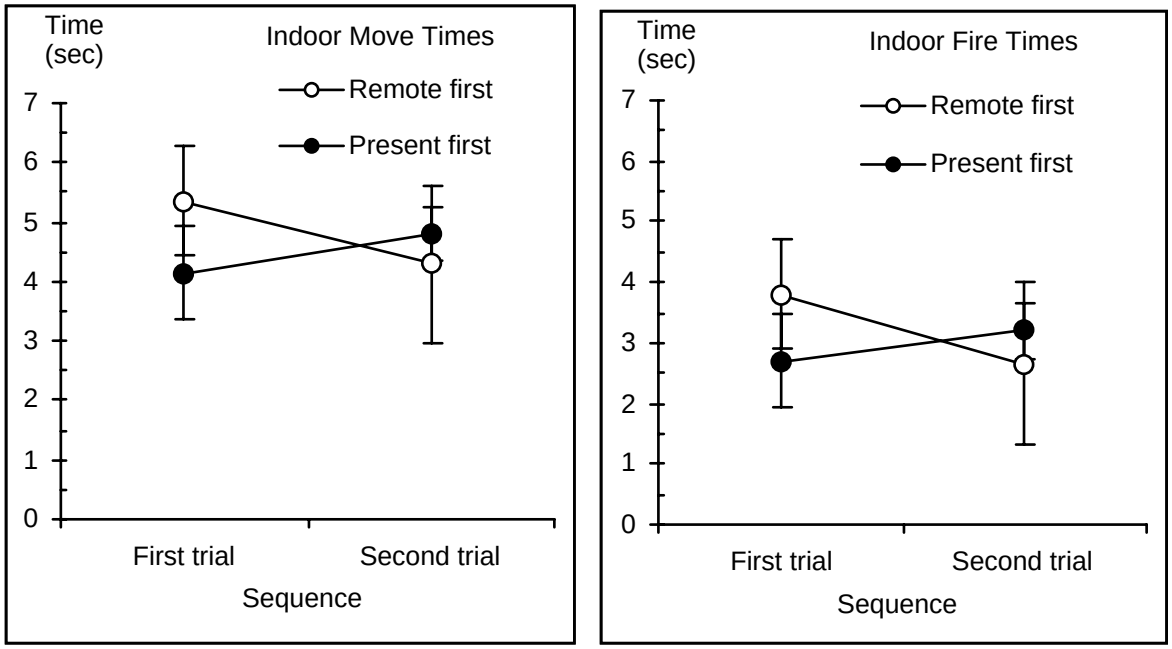


Figure 2 Data from Experiment 1 which was conducted inside a military building. A) Response times to the command to move. B) Response times to the command to fire. Responses are always faster in the leader-present condition.

Figure 2b is the corresponding graph of the response times to the command to fire. The pattern of results is the same: both groups of participants responded more quickly when the leader was present in the lane. A two-factor ANOVA found that neither group (remote-first, present-first), $F(1,36) = .120$, $\underline{p} > .73$, nor sequence (first trial, second trial), $F(1,36) = .155$, $\underline{p} > .69$, were significant. The within-subjects t-test for the effect of leader presence, $t(19) = 2.317$, $\underline{p} < .016$, indicates that leader presence made a significant difference in the participants' response times to commands to fire. The adjusted Cohen's d at an alpha of .05 is approximately .73. Again, the experiment had ample statistical power with 20 participants.

EXPERIMENT 2

Location and Participants

To test the generality of the indoor result from Fort Riley, we moved the second experiment to an outdoors venue on-campus. The setup was exactly the same as the lane in Figure 1 with one exception. The lane was set up in a small field rather than in a building. The change in setting made the leader-remote condition less remote. The first experiment was conducted indoors which allowed the remote leader to be completely out of sight. In the second experiment, the remote leader hid behind a tree approximately 50 meters behind the lane. Thus the leader was in fact visible if the participant chose to turn around and look. Twenty-two students, three women and 19 men, participated in the second experiment. The median age was 19 with a range of 18 to 26.

Results

The graphs of Figure 3 show the response times to the commands to move and to shoot. The open symbols show the means and standard errors of response times for the remote-first group. This group responded more quickly to both commands in the second trial when the leader was present in the lane. The closed circles, for the present-first group, show that participants responded more quickly in the first trial, again when the leader was present in the lane. The ANOVA on sequence

and group effects for the command to move show that group was significant, $F(1,40) = 3.779$, $\underline{p} < .058$. The remote-first group moved significantly more quickly in the second trial when the leader was present in the lane. The test for the effect of sequence, $F(1,40) = 1.17$, $p > .18$, shows no effect of sequence on move time. The ANOVA for fire time indicates that neither group nor sequence were significant, $F(1,40) = .284$, $p > .59$ and $F(1,40) = .181$, $p > .67$, respectively.

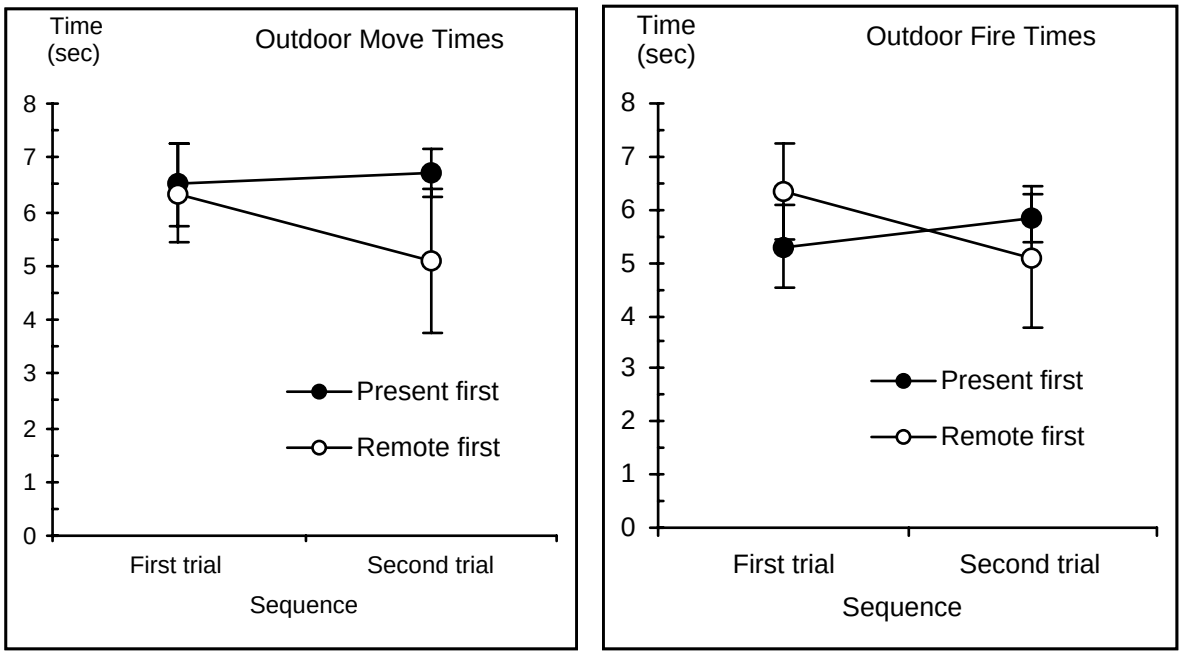


Figure 3 Data from Experiment 2 which was conducted outdoors. A) Response times to the command to move. B) Response times to the command to fire. Responses are always faster in the leader-present condition.

We merged the data across groups to conduct a within-subjects t-tests for the effect of leader presence. The test for both move times, $t(21) = 2.798$, $\underline{p} < .005$, and fire times, $t(21) = 2.211$, $\underline{p} < .019$, indicates that, as expected, leader presence made a significant difference in the participants' response times to commands to move and shoot. The adjusted Cohen's d at an alpha of .05 is approximately .55 for commands to move and .70 for commands to fire. The experiment had ample statistical power with 22 participants.

Given the similarity of the two experiments' results, it appears the subtle difference in the degree of remoteness of the leader across the two experiments did not have a significant impact on response times. The similarity also allows us to aggregate the data. The test on the composite move data is significant $t(41) = 4.122$, $\underline{p} < .0001$. The test on the composite shoot data is also significant $t(41) = 3.218$, $\underline{p} < .0013$. The aggregate power is very high.

EXPERIMENT 3

The third experiment is in progress at a commercial, indoor paintball arena in Tidan, near Skövde, Sweden. Most of the participants are students at the Skövde Högskolan (college). There is little reason to expect we will find a significant difference between Swedish youth and American youth when asked to follow commands to move and to shoot. The change in setting does, however, present the opportunity to address cross-cultural phenomena (Sutton, 2003). We are currently planning experiments to assess the effects of mixing leader and fire-team nationality.

The third experiment has added an intermediate condition (leader-radio-present) to disambiguate the effects of leader presence and mode of communication. In the intermediate condition, the leader is on the lane one station behind the participant communicating by radio. If leader presence is the major source of variability observed in the first two experiments, then performance in the leader-radio-present condition will be approximately the same as it is in the

leader-present condition. In contrast, if the effect is due to radio communication, then performance in the leader-radio-present condition will be like that in the remote-leader condition.

Biometric telemetry is being used to improve measurement of response time. Goniometers (strain gauges) are attached to the participants' and the leader's trigger fingers. Moving the finger stretches the gauge which changes the resistance that is telemetered to a portable computer (Biopac Systems MP150 system, with 2 TEL 100 C-RF remote monitoring modules). The leader bends his finger when he issues a command. Shooting and picking up new ammunition produce distinctive signals. Response times are calculated from the difference in the times of signals in the leader's and the participant's telemetered goniometer data. The telemetry system enables continuous electrocardiographic monitoring of the leader and selected participants. The resulting time series of interbeat intervals are the raw data for studying the correlation between experimental conditions and heart rate and heart-rate variability, two psychophysiologic measures of stress (Backs & Boucsein, 2000).

Data collection will be completed in February 2004 and the results reported at the conference.

DISCUSSION

These results from Experiments 1 and 2 support our hypotheses. Participants were faster to react to the leader's commands when the leader was present than when the leader was remote. This result suggests that some kind of intervention will be required if soldier performance is to be as efficient in remote command and control as it is in the more traditional, leader-present, mode of control. Two classes of intervention come to mind. The first is training. Does current military training overcome the inherent disadvantage posed by a leader's absence? We hope to address this question by conducting similar experiments with conscripts from the military garrisons in Skövde.

The second intervention is the development of technology that enables 'virtual leaders' to take to the field with their fire teams. The requirements for a virtual leader are not physical or holographic presence but psychological presence. We plan to test alternative designs for information telemetry and display that offset the decrements in performance that accompany remote command and control.

ACKNOWLEDGEMENTS

The views expressed in this work are those of the author and do not necessarily reflect official Army policy. This work was supported by the DOD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Dr. Elmar Schmeisser is the project monitor. The paintball paradigm and the assault lane were designed by Lt. Keith Pangburn.

REFERENCES

Milgram, S. (1974). Obedience to Authority. New York: Harper & Row.

Pangburn, K., Freund, S., Pangburn, H., & Smith, K. (2003). Beyond Milgram: An experimental study of leader presence. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society, 364-368.

Sutton, J. L. (2003). Validation of cultural awareness training concept. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society.

International Journal of Applied Aviation Studies, 4, 1
Copyright © 2004, FAA Academy, Oklahoma City, OK

Task Demand and Response Error in a Simulated Air Traffic Control Task: Implications for Ab Initio Training

L. L. Murphy
Department of Psychology, Institute of Simulation and Training, University of Central Florida

K. Smith
Department of Mechanical Engineering, Linköping Institute of Technology

and

P. A. Hancock
Department of Psychology, Institute of Simulation and Training, University of Central Florida

Abstract

This study investigated the relationship between task demand and the occurrence of error in an experimental simulation, which represented dynamic En-Route air traffic control. Participants were trained to baseline performance in the air-traffic-control task and then were presented with a series of 12 challenging but realistic scenarios. These scenarios were scripted to create two cycles of three levels of task demand as represented by traffic count. Conflict opportunities were scripted into each level of traffic count of six conflicts per scenario. Errors of omission were found to be equally likely when traffic count decreased from a peak as during a peak itself. This empirical finding was consistent with real-world experiences as reported in testimonial accounts by professional air traffic controllers. Given the restricted number of participants and their relative inexperience, we consider the present work to be an initial window into a highly complex issue. Our present findings have implications for ab initio training of air traffic controllers and also relate to performance in all operational domains, which demand flawless response from process control agents.

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-500-OU, P.O. Box 25082, Oklahoma City, OK 73125.

Introduction

The evolving complexity of human-machine systems has served to increase the demands placed upon its operator's limited capacity to process information (see Rochester & Komos, 1976; Smolensky & Stein, 1998). The fluctuating profile of task demand posed in dynamic environments can be expected to elicit a spectrum of behavioral, subjective, and physiological responses, which contribute to operator workload and performance (see Hancock, 1997; Melton, 1982). The increasing trend toward ever more complex technologies, which tax the human information processing system, makes it crucial to develop a thorough understanding of the relationship between task demand, the operator's response to that demand, and the subsequent outcome reflected in the on-going level of performance efficiency. The present study investigated these relationships through the specific and chosen use of a simulated air traffic control (ATC) environment, which permitted the careful quantitative manipulation of task demand level through control of the number of aircraft to be monitored. Since air traffic control represents a dynamic process in which there are explicitly defined operational errors (Greene, Muir, James, Gradwell & Greene, 1997; Hancock, 1997; Hopkin, 1995; Metzger & Parasuraman, 2001), it offers the opportunity to collect a wealth of response data making it an ideal setting for the empirical investigation of task demand and cognitive workload in a success-critical environment.

One of the significant and continuing problems in human performance assessment concerns the meaning of the term *workload*. Some researchers use the term workload to represent objective characteristics of externally based tasks, while others use the same term to represent the subjective experience of the individual placed in that environment (cf., Hancock & Meshkati, 1988; Stager & Hameluck, 1990; Stager, Hameluck & Jublis, 1989). It is evident that this usage as a property of both the operator and the operator's environment has led to a debilitating confusion. We here distinguish between workload as task demand (the external properties of the task independent of any individual) and the operator's response to that task demand, which is assuredly contingent upon the capabilities and skills of the individual so exposed. We suggest that the construct of workload is best defined as a subjective subset of an individual's response to task demand, being closely aligned with the notion of adaptation advanced by Hancock and Warm (1989). Given our representation of workload as a response of the exposed individual, we define *task demand* as a property of the environment itself. Clearly, in this division, the operator's workload response is contingent upon his or her appraisal of any immediate environment as presenting a *task* in the first place. Where such appraisal does not result in the perception of an immediate task to be resolved, no associated workload is experienced (Hancock & Chignell, 1988). Often this specification of a *task* is a result of a third-party arbiter who dictates exactly what that *task* is (Smith & Hancock, 1995). In this present experiment, we, the experimental team, imposed the task. In the real world, tasks arise from a myriad of necessities. In ATC, the controllers' task derives from the ultimate need to provide safe and efficient air transportation.

In a constantly changing task environment, there are often many different sources of information. This profile of information is rarely exactly repeated and for each situation is arguably unique (Hancock, Flach, Caird & Vicente, 1995; Hancock & Meshkati, 1988). Task demand is crucially dependent on the type and amount of task-relevant information presented, and information changes the operator's behavior. We illustrate this by examining the task demand in simulated air traffic control where operators seek to maintain at least the minimum legal separation between aircraft (see Smith, Scallen, Knecht & Hancock, 1998). The level of task demand is a complex function of not only the assimilation and evaluation of task-relevant information, but also of the need to execute a number of cognitive and physical behaviors. The cognitive behaviors include recognizing sources of information on both analog and alphanumeric displays and deciding to intervene (or not) by issuing commands to the pilot (Smith & Murphy, 2000; Smolensky & Stein, 1998). The physical behaviors include monitoring the display, verbally issuing commands to the pilot, handing off aircraft to adjacent sectors, and organizing flight strips (Hopkin, 1995). Sources of task-relevant information include display symbols for aircraft, their vectors and intent, flight strip identifiers for aircraft and their flight-plans, and verbal and written input from traffic management coordinators, from the national weather service, from adjacent sector controllers, and from pilots themselves (Smith & Mafera, 2000). In our simulation-experiment, all sources of information other than traffic count were held constant. The independent variable in the experimental scenarios - the increasing, peaking, and decreasing of traffic count – was therefore the only manipulation of task demand in our study.

We define *response to task demand* as the ensemble of mental actions, overt behaviors, and physiological responses that follow from the operator's interaction with his or her task environment in order to fulfill the goal of the task (Hancock & Desmond, 2001; Melton, McKenzie, Polis, Hoffman, & Saldivar, 1973; Melton, McKenzie, Polis, Funkhouser & Lampietro, 1971; Thackray, Bailey, & Touchstone, 1975; Wilson & Corlett, 1999). Both acceptable performance and error are part of the operator's response to task demand. Another natural measure is task-relevant communication. (Carlson, 1982; Chapanis, 1953; Hendy, 1998) Our measures, the occurrence of error and of task-relevant communication, are overt behavioral indicators of the operator's response to the levels of task demand posed by the experiment.

Task Demand and the Occurrence of Errors

Errors are the nemesis of process control agents working in a system like air traffic control that has a high potential for risk (Smith, Briggs, & Hancock, 1997). *Operational errors* occur whenever two aircraft under positive control violate each other's protected zone (a violation of the minimum separation) which is a compound criterion of 5 miles longitudinally and 1000 feet vertically (Rodgers & Nye, 1993). These formal operational errors are distinguished from the *true errors* of omission and/or of commission that necessarily precede them. Operational errors are viewed not as errors *per se* but as the product of true error at some

earlier point in time. Some true errors are errors of omission: the controller failed to take action that would maintain separation. Other true errors are errors of commission: the controller instructed an aircraft to make a maneuver that directly led to an operational error.

Several previous studies have examined controllers' subjective appraisal of workload (light, moderate, or high) and the occurrence of operational errors (Arad, 1964; Kinney, Spahn & Amato, 1977; Redding, 1992; Rodgers & Nye, 1993; Rodgers, Mogford, & Mogford, 1998; Schmidt, 1976; Stager et al., 1989; Stager & Hameluck, 1990). A review of these relevant studies reveals two themes. First, operational errors tend to occur when traffic count is low to moderate (approximately eight aircraft in the sector). Second, when controllers are asked to subjectively rate traffic volume and workload at the time of the operational error, they also tend to rate it as low to moderate.

It is unclear; however, whether these findings reflect (1) a decreased tendency to make operational errors under high traffic conditions or (2) the lower frequency of high traffic conditions in general (Endsley & Rodgers, 1997). Additionally, it is unclear from these reports whether the true error (of omission or of commission or a combination of these two influences) occurred (A) in low and moderate traffic conditions or (B) earlier when the level of traffic was relatively high. In either case, it may have manifested as an operational error only later when traffic had decreased to a low or moderate level. This source of uncertainty in linking the spatio-temporal occurrence in true error to reported operational errors is a major concern, which we have termed *the phase delay dilemma*.

The Phase Delay Dilemma

Phase-delayed errors are common in dynamic environments (Smith et al., 1998). In the ATC domain, the time lag between a true error and the operational error poses a fundamental dilemma: How are we to know when the true error actually occurred? For example, suppose an operational error occurred during a peak in traffic count. Did the true error also occur during this peak or did it occur earlier as the level of traffic was increasing? Due to the phase delay dilemma, previous research into operational errors may not reveal the actual link between true errors and operational errors and between true errors and task demand. It is therefore not reasonable to conclude that an operational error is a function of a particular level of traffic when the relationship between the true errors and the level of task demand still remains uncertain.

Our experiment was designed to address this impasse by scripting air traffic scenarios that controlled for the phase delay dilemma and by obtaining behavioral measures of performance and response to task demand. Our premise is that systematically changing traffic count in realistic En Route air traffic scenarios result in systematic variations in task demand. By creating two peaks of traffic count in a scenario, we control for the potential confound of operator fatigue. By incorporating conflict opportunities within each level of traffic count,

we also address the potential effects of phase-delayed operational errors and make it possible to distinguish between errors of omission and errors of commission.

Experimental Method

Experimental Participants

Three, upper level, undergraduate students volunteered to act as the air traffic controllers over a six-week period. A second trio of students acted as pseudo-pilots. Participants received class credit for their participation. All participants received 6 practice and 12 experimental scenarios over the six-week period. Scenarios were counterbalanced using a selection from a Latin Square. Students were used rather than professional air traffic controllers for two reasons. First, the purpose of the study was to observe and document errors and inexperienced participants are much more likely to make errors than professionals are. Second, recent national security events have made it extremely difficult to work with professional air traffic controllers (Hancock & Hart, 2002), although we would have preferred to do so. We are fully aware that inexperienced participants are inclined to make different errors than experienced participants (Reason, 1990). Therefore, our results may generalize most to novices like those undergoing ab initio training. However, in mitigation of such issues, at the present stage we are more concerned with contributing to basic empirical understanding than with the immediate domain-specific application of such knowledge.

Experimental Task

The student-controllers were responsible for maintaining the Federal Aviation Administration's criterion for minimum separation between aircraft by issuing appropriate verbal commands to the pseudo-pilots. Command options included changing an airplane's heading, altitude, and/or speed. Pseudo-pilots provided verbal confirmation of controller commands and maneuvered their aircraft accordingly. Pseudo-pilots were trained along side the controllers and were instructed to complete the controller's commands as quickly and as accurately as possible. Analysis of scenario histories confirmed their accuracies in responding, which leads us to conclude that this element of the simulation did not influence subsequent results. During the two weeks of the practice session, the students learned how to use the experimental platform. They became adept at monitoring dynamic traffic, at identifying and resolving potential operational errors, and at giving and confirming verbal commands designed to maintain separation. The criterion for baseline performance was resolution of six scripted aircraft conflicts in a 30-minute practice scenario.

Experimental Platform

The experiments were run using the Distributed Air Traffic Information Display Simulator (DATIDS), a full simulation of an ARTCC sector controller's workstation (Klinge, Smith, & Hancock, 1997). The simulator presents a representation of the composite radar screen, the computer read-out display, and the but-

tons and dials used to adjust the settings of the R-side ATC display and an illustration of this simulation is shown in Figure 1.

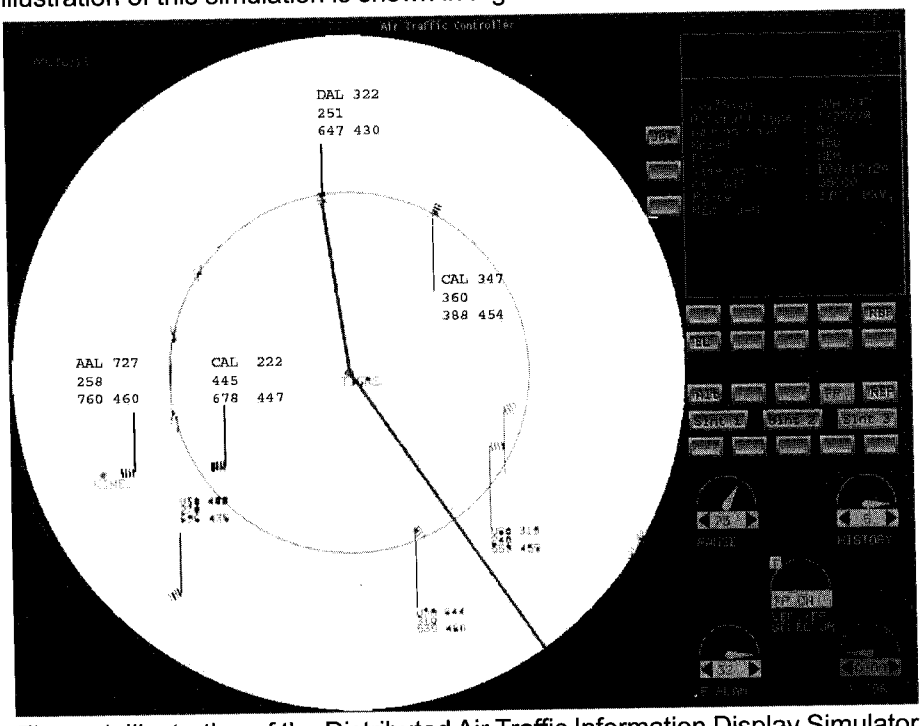


Figure 1. Illustration of the Distributed Air Traffic Information Display Simulator (DATIDS).

Experimental Design

As shown in Figure 2, the experimental scenarios incorporate two blocks of time, each containing a cyclic pattern of traffic count with three levels (5 minutes of increasing traffic, 5 minutes of peak traffic, and 5 minutes of decreasing traffic). The two 15-minute blocks and three levels of traffic count yield a 2x3 repeated measures design. The repeated blocks permit an assessment of the potential confound of fatigue. Task demand is operationally defined as the average traffic count, being the average number of aircraft visible on the controller's information display at each minute. More specifically, in the first period of five minutes of each block, the traffic count continually increased from three aircraft to a peak value of 16 aircraft. During the second period of five minutes of each block, the traffic count remained at or near this peak. In the third period of five minutes of each block, the traffic count gradually decreased until it returned to the baseline of three aircraft. These criteria, of a minimum of three aircraft and a maximum of 16 aircraft, were adopted based on actual observations of a controller's typical sector load (Smith & Murphy, 2000).

Controlling for Phase Delay

To make errors of omission evident, two aircraft were scripted to create two

distinct opportunities for an error of omission in each of the six levels of traffic count. The two aircraft entered and exited the information display during the same (five minute long) level of traffic count. If the student-controller took no action to maneuver these aircraft, then the resulting operational error could be attributed to an error of omission that occurred during that level of traffic and not in a prior level. This control makes it possible to know the number of aircraft and the level of task demand when and if errors of omission occurred. All other aircraft were scripted to be conflict-free. Therefore, if any other errors occurred (at any time) they could be attributed to an error of commission.

Dependent Measures

The primary performance measure was the number of operational errors due to errors of omission or due to errors of commission occurring throughout the experiment. The behavioral measure of response to task demand was the number of commands between controller and the pseudo-pilot.

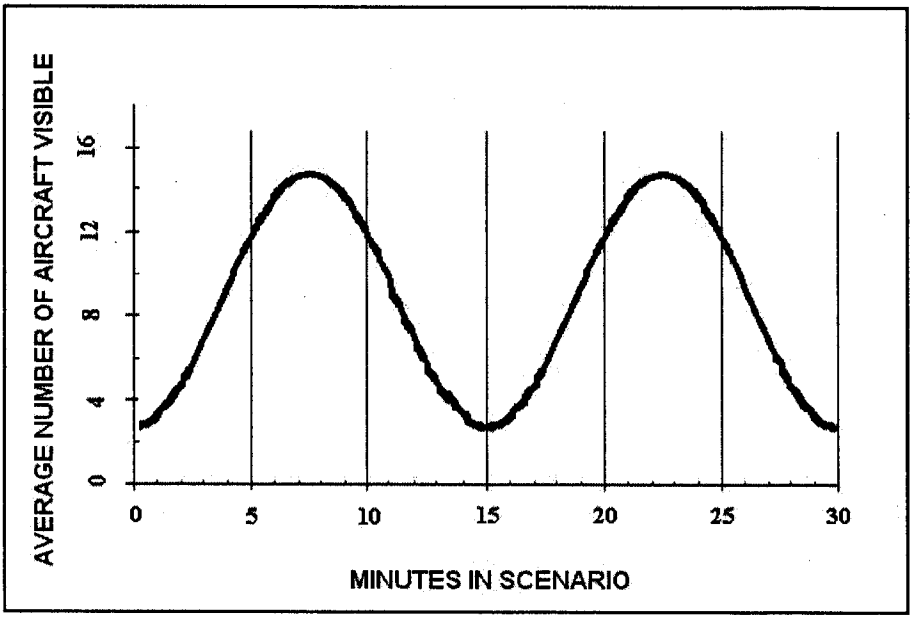


Figure 2. Illustration showing the schematic design of the experimental scenarios.

The horizontal axis is elapsed time in the scenario. The vertical axis is the minute-by-minute average number of aircraft visible on the controller's information display. Each scenario lasted 30 minutes and presented two cycles of traffic count. Each cycle contained three periods of five-minute-long conditions of traffic count: increasing, peak, and decreasing. In each cycle the average number of aircraft increased for a period of five minutes from a baseline of approximately three aircraft, remained near a peak level of approximately 16 aircraft for five minutes, and then decreased for a period of five minutes back to the

base level. Opportunities for operational errors due to errors of omission were scripted into all six periods of five-minute-long conditions of traffic count.

Experimental Results

A repeated measures 2x3 (two blocks of time by three levels of traffic count) analysis of variance was used to analyze the error and communication data. The number of scenarios (12) acted as a proxy for number of participants in order to increase the statistical power of the test. Mauchly's test of sphericity was not significant for any of the analyses and thus sphericity was uniformly assumed.

Errors of Omission

Figure 3 shows the total number of errors of *omission* for each level of traffic count (increasing, peak, and decreasing) and both blocks of time. A 2x3 repeated measures ANOVA was conducted to determine the effect of traffic level and block of time on the number of operational errors due to errors of omission. The main effect for traffic level was found to be significant, $F(2, 22)=10.73, p<.001$. A within-subjects $eta^2=.50$ indicates a medium effect size for traffic level. Post-hoc pair-wise comparisons revealed no significant difference between the mean number of errors of omission in the peak and decreasing conditions. However, both the peak and decreasing conditions were significantly different from the increasing condition for both blocks of time, $p<.05$. It appears that the student-controllers were just as likely *not* to take corrective action to resolve potential conflicts after a peak in traffic as they were during a peak in traffic for both blocks of time in the scenario.

Errors of Commission

Figure 4 shows the total number of errors of *commission* for each level of traffic count and both blocks of time. A 2x3 repeated measures ANOVA was conducted to determine the effect of traffic count (increasing, peak, and decreasing) and block of time on errors of commission. Traffic level was not found to be significant. However, there was a significant difference for blocks of time, $F(1, 11)=7.86, p<.01$. The within-subjects $eta^2 = .42$ indicates a medium effect size. While many more errors of commission were made in the second block than in the first block, the totals are low but remain significant. Errors of commission were relatively rare. They occurred when student-controllers issued inappropriate commands that caused an operational error between aircraft that otherwise would have remained separated. These results suggest that the student-controllers experienced some amount of fatigue or vigilance decrement (Mackworth, 1948; Mackworth, 1957; Hancock, 1984) or both.

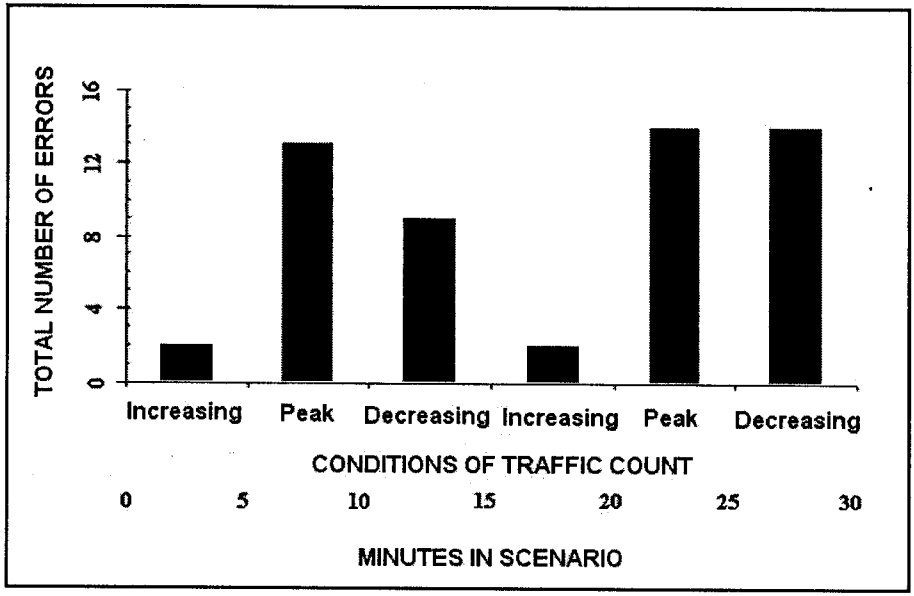


Figure 3. Total number of operational errors due to errors of *omission* collapsed across scenarios and conditions.

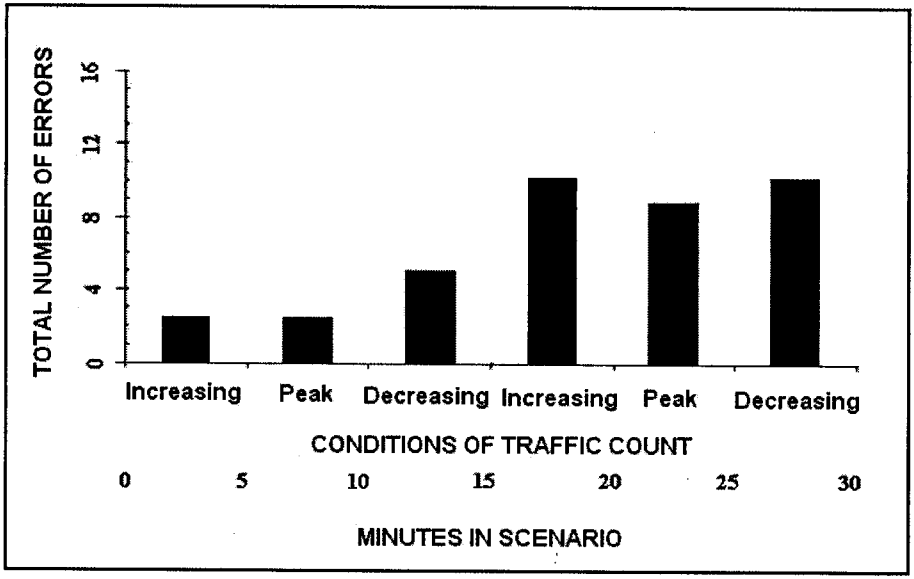


Figure 4. Total number of operational errors due to errors of *commission* collapsed across scenarios and conditions. The vertical scale is the same as in Figure 3 to facilitate comparison.

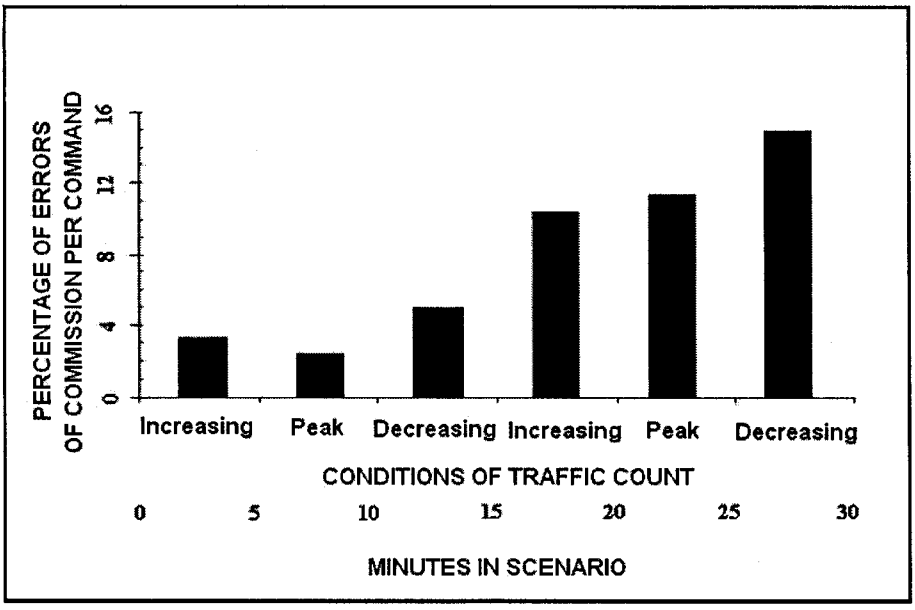


Figure 5. Percentage of errors of commission per command to maneuver aircraft issued by participants collapsed across scenarios and conditions.

Controller Communication

Figure 5 shows the percentage of errors of commission per command to aircraft issued by the student-controllers for each level of traffic count and both blocks of time. A 2x3 repeated measures ANOVA was conducted to determine the effect of traffic level and block of time on the communication. The interaction between the three levels of traffic and the two blocks of time was significant, $F(2, 22)=8.29$, $p<.002$. The within-subjects $eta^2 = .43$ indicates a medium effect size. The interaction suggests that the relationship between communication and errors of commission is such that more errors of commission are committed per command during the second block of time. This may be due to fatigue, diminished attention capacity, or both.

Correlations

A correlation analysis was conducted to detect emergent relationships between the dependent and independent variables. The analysis, shown in Table 1, found significant positive correlations between two of the three dependent variables, errors of omission and communication, and the experimental treatment, level of traffic count. These results indicate that the manipulation of traffic count had an effect on participant behavior. Both types of error were found to correlate significantly with communication. The correlation between communication and errors of commission was positive. This result suggests that errors of commission tended to occur during periods when the student-controllers' response to task demand was relatively high. In contrast, the correlation between communication and errors of omission was negative. This result suggests that

inadvertent operational errors tended to occur when the student-controllers' response to task demand was relatively low.

Table 1
Correlations between the three dependent variables - errors of omission, errors of commission, and communication – and the main dependent variable, traffic count.

	Omission	Commission	Communication	Traffic
Omission	1.00	0.03	-0.14 *	0.21 **
Commission		1.00	0.22 **	-0.06
Communication			1.00	0.13 *
Traffic				1.00

* Correlation is significant at the .05 level (2-tailed)
** Correlation is significant at the .01 level (2-tailed)

Discussion

The purpose of this study was to develop an understanding of the relationships between task demand, the operator's behavioral response to that demand, and the occurrence of errors of omission and of commission in a simulated ATC environment. The data suggest that errors of omission are equally likely to occur after a peak in task demand as during that peak and are associated with relatively low levels of response to task demand. Errors of commission are more likely to occur when response to task demand is unusually high and time-on-task has made it likely that fatigue has set in. In short, this type of error is sensitive to fatigue, to the level of task demand, and to the operator's behavioral response to that task demand. Previous investigations of task demand, response to task demand, and the occurrence of errors in the ATC domain (e.g., Kinney et al., 1977; Stager, Hameluck, & Jublis, 1989; Stager & Hameluck, 1990) has focused on workload, the operator's subjective response to task demand. Using archival data, they reported that operational errors were more likely to occur when workload was said to be moderate to low. Since the data were archival, these studies failed to control for the phase delay dilemma, the time lag between the operational error and the error of omission or commission that necessarily preceded them.

Errors of Omission

The total number of errors of omission across conditions shown in Figure 3 revealed that operational errors due to errors of omission were equally likely to occur after a peak in traffic as during a peak. This finding was consistent with anecdotal evidence reported by professional air traffic controllers. Most controllers freely admit that operational errors occur most often on the backside of a

peak in traffic. Unlike previous archival studies investigating operational errors, the current study scripted the simulated traffic scenarios to control for traffic phase-delay. Consequently, the current study contributed a new finding: operational errors occur equally as often on the backside of the peak in traffic as they do during the peak. These results were found for both blocks of time, beginning of scenario (first 15 min) and end of scenario (last 15 min), suggesting that operational errors due to errors of omission are more likely a function of task demand than of fatigue. If the effect were due to fatigue following high traffic levels, then significantly more errors would have occurred at the end of the scenarios, which was not the case. These findings suggested that the decrease of task demand after a peak is just as taxing on inexperienced participants as the peak itself. These participants were expending relatively high levels of effort as they made errors of omission. Operational errors due to omission were not often a result of the inexperienced participant dropping his or her guard.

The correlation between errors of omission and communication was negative in this experiment, see Table 1. Our inexperienced participants issued inappropriate commands to aircraft while missing critical cues about impending conflicts. It appears that the operational errors due to errors of omission in this experiment may reflect our participants' relative lack of skill. Accordingly, these present results may generalize only to inexperienced students in Collegiate Training Initiative institutions that are beginning their training to become air traffic controllers. Whether such results generalize to actual operations or other operations in other process control domains requires further empirical evaluation.

Errors of Commission

The number of errors of commission steadily increased as scenarios progressed, see Figure 4. This result suggested that operational errors due to errors of commission are more a function of fatigue than of the cyclic manipulation of task demand. Additional evidence of a fatigue effect was found in the statistically significant pattern of the errors of commission per command, shown in Figure 5. These findings were congruent with the vigilance literature, which reported that the frequency of correct detections tends to decrease after 20 minutes on watch (e.g., Hancock & Warm, 1989; Mackworth, 1948; Mackworth, 1957; Parasuraman, 1986; Warm, 1984). This deterioration in performance is traditionally termed the "vigilance decrement."

Limitations of the Study and Future Directions

The correlation data in Table 1 show that the cyclic manipulation of traffic was positively correlated with two of the dependent measures, communication and errors of omission. An analysis of variance found that the fluctuation of task demand had an effect on the participants' responses to task demand. These results suggested that this study had high internal validity. However, there are two significant threats to the external validity of this study. The first is nature and the small size of our sample – three participants, all of whom were undergradu-

ate psychology students. However, these students were trained and participated in the study for over six weeks (two weeks of the practice session and four weeks of the experimental session). Therefore, the results may generalize primarily to inexperienced trainees striving to be air traffic controllers. Future studies should of course address a larger group of participants from a pool that is more representative of trainees. That said, it must be noted that Federal and union regulations make it extremely difficult to gain access to these FAA trainees. Eventually if the work is to exert a strong practical impact, it might be replicated and extended with full-time professional controllers. The second threat is construct validity. In debrief sessions none of the student-controllers mentioned the cyclic pattern of task demand and the relatively regular occurrence of the scripted conflicts. Nevertheless, it is possible that they were implicitly aware of them, made tacit hypotheses about them, and responded to those hypotheses. Construct validity may be improved by counterbalancing the order of the presentation of conditions (e.g. from {increase, peak, decrease} to {peak, decrease, increase} etc.), and by adding control conditions with uniform levels of task demand in order to get baselines of performance (see Hancock, Williams, Manning, & Miyake, 1995).

The major finding in this study was that errors of omission are equally as likely on the backside of a peak in task demand as during the peak. It would be useful to determine whether this finding generalizes to other process control and continuous operator tasks and other dynamic environments. Tasks and environments that are amenable to laboratory simulation include operating a motor vehicle, piloting an aircraft, firing weapons, and operating an assembly line. Such studies would be especially effective if the order of the presentation of levels of task demand was fully counterbalanced and control conditions used to get baselines of performance. For example, one useful control condition would pose scenarios with two cycles of traffic demand but no scripted conflicts. Other control conditions would pose scenarios with constant (but different) levels of traffic demand and several scripted conflicts. The resulting within-subject comparisons would provide a strong test of the generality of our results. If this preliminary finding of errors of omission on the backside of peaks in task demand holds up to further testing, it would become imperative to investigate ways of mitigating this effect. Eye tracking during the task could provide detailed information about the information the operator is focusing on (hits vs. noise) and might shed light on these sources of errors and, in turn, on ways these errors might be avoided.

Acknowledgements

We would like to thank Scott Shappell and William Knecht for their critiques of an earlier version of this paper. We also are grateful for the most helpful comments of the three anonymous reviewers in revising our work.

References

Arad, B.A. (1964). *Notes on the measurement of control load and sector design in the en-route environment.* (Report No. AD-659035). Washington, D.C.: Federal Aviation Administration.

Carlson, G.E. (1982). *Captain Carlson's airplane talk: The complete book of VFR and IFR communications.* Las Vegas: Watosh Publishing.

Chapanis, A. (1953). *Research techniques in human engineering.* Baltimore: Johns Hopkins Press.

Endsley, M.R. & Rodgers, M.D. (1997). *Distribution of attention, situation, awareness, and workload in passive air traffic control tasks: Implications for operational errors and automation.* (Report No. DAOT/FAA/AM-97/13). Washington, D.C.: Federal Aviation Administration.

Green, R.G., Muir, H., James, M. Gradwell, D. & Greene, L.G. (1997). *Human factors for pilots*. Vermont: Ashgate Publishing Co.

Hancock, P.A. (1984). Environment stressors. In: J.S. Warm (Eds.). *Sustained Attention in Human Performance.* (pp. 103-142), New York: Wiley.

Hancock, P.A. (1997). *Essays on the future of human-machine systems*. MN: Banta Information Services.

Hancock, P.A., & Chignell, M.H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man, and Cybernetics*, 18, 647-658.

Hancock, P.A. & Desmond, P.A. (2001) (Eds.). *Stress, workload, and fatigue: Human factors in transportation*. Mahwah, New Jersey: Lawrence Erlbaum Associate Publishers.

Hancock, P.A., Flach, J., Caird, J., & Vicente K. (1995). (Eds.). *Local applications of the ecological approach to human-machine systems*. Mahwah, New Jersey: Lawrence Erlbaum Associates Inc.

Hancock, P.A. & Hart, S.G. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in Design, 10 (1)*, 6-16.

Hancock, P.A. & Meshkati N. (Eds.) (1988). *Human mental workload*. Amsterdam: North-Holland.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519-537.

Hancock, P.A., Williams, G., Miyake, S., & Manning, C.M. (1995). The influence of task demand characteristics on workload and performance. *International Journal of Aviation Psychology, 5*, 63-85.

Hendy, K C. (1998). CRM: More than just talk, talk, talk...*Flight Safety Spotlight, 98(1)*, 12-15.

Hopkin V.D. (1995). *Human factors in air traffic control*. England: Burgess Science Press.

Kinney, G.C., Spahn, J., & Amato, R.A. (1977). *The human element in air traffic control: Observations and analyses of the performance of controllers and*

supervisors in providing ATC separation services. (Report No. MTR-7655). McLean, VA: METREK Division of the MITRE Corporation.

Klinge, J., Smith, K., & Hancock, P.A. (1997). DATIDS: The University of Minnesota distributed air traffic information display simulator. *Proceedings of the 9th International Symposium on Aviation Psychology*, Columbus, Ohio.

Mackworth, N.H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology, 1*, 6-21.

Mackworth, N.H. (1957). Some factors affecting vigilance. *Advancements in Science, 53*, 389-393.

Melton, C.E, McKenzie, J.M., Polis B.D., Funkhouser, G.E & Iampietro, P.F. (1971). *Physiological responses in air traffic control personnel: O'Hare Tower.* (Report No. DOT/FAA/AM-71-2). Washington, D.C.: Federal Aviation Administration.

Melton, C.E, McKenzie, J.M., Polis B.D., Hoffman, M., & Saldivar, J.T. (1973) *Physiological responses in air traffic control personnel: Houston Intercontinental Tower* (Report No. DOT/FAA/AM-93-22). Washington, D.C.: Federal Aviation Administration.

Melton, C. E. (1982). *Physiological stress in air traffic controllers: A review.* (Report No. OAM/FAA/AM-82/17). Washington, D.C.: Federal Aviation Administration.

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors, 43(4)*, 519-528.

Parasuraman, R. (1986). Vigilance, monitoring, and search in K. R. Boff L. Kaufman, and J. P. Thomas (Eds.), *Handbook of Perception and Human Performance: Cognitive Processes and Performance* (Vols. 2.). New York: Wiley.

Reason, J.T. (1990). *Human error.* Cambridge, UK: Cambridge University Press.

Redding, R.E. (1992) Analysis of operational errors and workload in air traffic control. *Proceedings Human Factors Society, 36,* 1321-1325.

Rochester, S.I. & Komos, N.A. (1976). *Takeoff at mid century. Federal Aviation policy in the Eisenhower years 1953-1961.* Washington, D.C.: Federal Aviation Administration.

Rodgers, M.D., & Nye, L.G. (1993). Factors associated with the severity of operational errors at air route traffic control centers. In M.D. Rodgers (Ed.), *An examination of the operational error database for air route traffic control centers* (Report No. DOT/FAA/AM-TN93/22, NTIS No. ADA275986). Washington, D.C.: Federal Aviation Administration.

Rodgers, M.D., Mogford, R.H., & Mogford, L.S. (1998). *The relationship of sector characteristics to operational errors* (Report No. DOT/FAA/AM-98/14). Washington, D.C.: Federal Aviation Administration.

Schmidt, D. K. (1976). On modeling ATC workload and sector capacity. *Journal of Aircraft, 13(7)*, 531-7.

Smith, K., Briggs, A., & Hancock, P.A. (1997). Success and failure at self-separation in simulated free flight. *Proceedings Meeting of the Human Factors and Ergonomics Society, 41,* 13-17.

Smith, K., & Hancock, P.A. (1995). Situation awareness is adaptive, externally-directed consciousness. *Human Factors*, 37, 137-148.

Smith, K., & Mafera, P. (2000). *ATC coordinator's information requirements for the NAS.* (Report No. AAR-100). Washington DC: The Office of the Chief Scientist and Technical Officer for Human Factors of the Federal Aviation Administration.

Smith, K., & Murphy, L. (2000). *Traffic management unit structure, positions, and uses of the traffic situation display.* (Report No. AAR-100). Washington DC: The Office of the Chief Scientist and Technical Officer for Human Factors of the Federal Aviation Administration.

Smith, K. Scallen, S.F., Knecht, W., & Hancock, P.A. (1998). An index of dynamic density. *Human Factors, 40*, 69-78.

Smolensky, M.W. & Stein, E. S. (1998). *Human factors in air traffic control.* California: Academic Press.

Stager, P., & Hameluck, D. (1990). Ergonomics in air traffic control. *Ergonomics, 33*, 436-9.

Stager, P., Hameluck, D., & Jublis, R. (1989). Underlying factors in air traffic control incidents. *Proceedings of the Human Factors Society, 33,* 43-46.

Thackray, R.I., Bailey, J.P., & Touchstone, R.M., (1975). *Physiological, subjective, and performance correlates of reported boredom and monotony while performing a simulated radar control task.* (Report No. DOT/FAA/AM-75-B) Washington, D.C.: Federal Aviation Administration.

Warm, J.S. (Ed.) (1984). *Sustained attention in human performance.* Chichester: Wiley.

Wilson, J.R. & Corlett, E.N. (1999). *Evaluation of human work: A practical ergonomics methodology*, (2nd ed.), Philadelphia: Taylor & Francis Inc.

Grazioli, S., Smith, K., & Johnson, P. J. (2004). Managing risk in social exchange. In K. Smith, J. Shanteau, & P. E. Johnson (Eds.), Psychological Explorations of Competent Decision Making (pp. 71-123). Cambridge, UK: Cambridge University Press.

3 Managing Risk in Social Exchange

Stefano Grazioli, Kip Smith, and Paul Johnson

In this chapter we present an analysis of decision making to manage risk in environments characterized by social exchange. We take a pragmatic approach, defining risk as the potential for loss and social exchange as an interaction where two or more agents define the outcomes (costs and benefits) for each other. We propose an information processing model of decision making to manage risk in such environments. The critical steps in the model are (1) ascribing beliefs and goals to the other agent(s) and (2) modifying behavior to reflect the ascription. We use the model to investigate the decision making of professionals who manage risk in two different environments characterized by social exchange: spot currency trading (banking) and second partner review of financial statements (auditing). We argue that ascription is an important component of competent decision making in both domains. The spot currency market is a dynamic environment that places a premium on real-time decision making. Traders ascribe beliefs to the market in order to anticipate the direction of prices. Financial statements auditing is characterized by the potential for opportunistic manipulation of information by its provider (management). Auditors ascribe intentions to management in order to detect these manipulations. To evaluate the role of ascription in risky social exchange, we (1) conducted experiments with professional subjects, (2) built two production systems, one for each domain, that implement the ascription process and emulate the behavior of the subjects in our experiments, (3) excised the ascription process from the systems to create a pair of alternative systems, and (4) compared the systems' and subjects' performance. The comparison demonstrates that ascription of beliefs and goals to others is a major determinant of success in managing risk in social exchange. This thinking about others' thinking is the crux of metacognition-others.

Risk in Social Exchange

During the past several years, we have been investigating decision making in two managerial domains: spot currency trading and second partner review of financial statements. Currency traders are bankers who manage the risk associated with exchanging currencies in markets characterized by fluctuating prices. These price fluctuations depend largely on the aggregate behavior of the other traders who participate in each market. Second partner reviewers are auditors who manage the risk associated with providing assurance that financial information used as a basis for decision making is reliable. The risk in this instance is that a client company intentionally manipulates its financial statements for the purpose of favorably affecting the behavior of the users of that information. For all their differences, these two domains share a defining characteristic: They both require agents to manage forms of risk that are dependent on the behavior of others.

We depart from more traditional accounts that equate risk with probability (e.g., Keynes, 1921; von Neumann & Morgenstern, 1944) and take a more pragmatic approach to define risk as the potential for loss (March & Shapira, 1987; Schoemaker, 1993; Shapira, 1995; Yates & Stone, 1992). Loss is generally defined as a negative change in an agent's economic position. Risk, being the opportunity for loss, does not necessarily generate a loss. Loss is the realization of risk. Risk is present whenever you stand to lose. Serious risk is present whenever you stand to lose a lot.

Social exchanges are situations where two agents interact to define outcomes – that is, costs and benefits – for each other (Cosmides, 1985; Cosmides & Tooby, 1992, 1995). As a result, one agent's action may negatively affect the other's outcome. Such a possibility for loss implies that risk is an inherent characteristic of social exchange. The possibility of losing millions of dollars in the blink of an eye by holding the wrong position in the foreign currency markets is an example of risk in social exchange. The potential for a lawsuit and loss of professional prestige that arises from failure to detect a financial fraud is another.

We argue that managing situations where others, acting alone or in concert with your behavior, provide the opportunity for loss is a critical aspect of decision making. Although domain knowledge is clearly necessary for most decision-making situations in the world, an additional kind of knowledge is necessary to support competence at decision making characterized by risks posed by others. This addition – the process

foreign currency. The word *market* refers both to the arena for exchange and to the network of traders actively trading.

The currency trader is a banker. Banks participate in the market in order to hold foreign currencies and to provide the service of exchanging currencies for customers. The trader's job is to manage the risk posed to the bank by participating in the currency markets. Because a currency is essentially worthless except in its country of issue and in the international markets, a foreign currency is a risky asset that is subject to the vagaries of international events. Traders refer to the asset they manage as a *position*. Position size is typically on the order of millions of dollars.

Unlike the futures markets for commodities, where trading is conducted in a *pit* (e.g., the Chicago Board of Trade), there is no face-to-face action in currency trading. Currency markets are electronic markets or *cybermarkets* (Kurtzman, 1993); they exist only in a web of specialized information systems. All information arrives at the trading desk across worldwide communications networks provided by wire services (e.g., Reuters, Bridge) or by telephone.

The trader samples five types of information: (1) an *overnight report*, a synopsis of global events prepared by the bank, (2) the prices currently showing in the market, that is, the exchange rates, (3) headlines from around the world, (4) the size and breakeven rates of his or her positions, and (5) the behavior of trading partners. Traders everywhere receive at approximately the same time essentially identical information about exchange rates and headlines. The only exceptions to the universality of information are the private data about positions and the in-house research done to support the trading function.

The currency markets are largely unregulated. Traders rely on the integrity of their trading partners and their bank. Both parties to a trade assume that the other's word is good. Each understands that the other is unwilling to assume unwarranted losses. Each anticipates that the other will quote rates close to the prices showing on his or her monitor. This mutual knowledge and the respect it engenders are built upon the scores of trades the trader makes each day. The countless iteration of one-on-one social exchanges supports cooperation and punishes deception (Holland, 1975/1992). Prices are the product of the cumulative interaction of this network of traders engaged in social exchanges. Fluctuations in prices reflect the fluid consensus of the network.

Traders make money by holding positions that most others in the market want. The value of these positions increases as the exchange

rates change. Traders lose money by holding positions that no one else wants. The need to predict exchange-rate fluctuations creates the need to understand and anticipate others' behavior. This need constrains the knowledge the trader brings to the task: The spot currency trader needs to know how to understand and anticipate the behavior of the market, that is, to understand and anticipate the behavior of a network of traders. Although all traders pursue their own self-interests, they cannot afford to forget that other traders know them as well as they know the others. Survival in the spot currency markets depends on one's ability to understand what the market wants.

Asymmetry of Information Under Conflict of Interest

Asymmetry of information is the second condition that engenders risk in social exchange. Information is asymmetrically distributed when somebody possesses information that others cannot readily access. The asymmetry makes those in need of information vulnerable to opportunistic behavior. When an agent relies on asymmetrically distributed information to make a decision, the provider of such information may opt to manipulate it for the purpose of steering the decision process toward a desired, favorable outcome.

From the perspective of the information user, risk is the possibility that the information used to make a decision has been manipulated by its provider. The realization of risk, that is, the actual occurrence of a successful manipulation, produces a loss because the decision is made on the basis of tainted information. The decision outcome is not what the decision maker would have selected without the manipulation.

Second partner review is an auditing task that exemplifies the risk posed by asymmetric distribution of information between the management of a company and the public of individuals who make decisions based on the financial statements issued by the company. Because many of these decisions (e.g., investing in the company's securities or granting it a loan) have serious consequences for the company's well-being, management is motivated to manipulate the reported financial information to influence the behavior of these decision makers (Johnson, Grazioli, & Jamal, 1993).

Independent auditors have the general goal of verifying that financial statements issued by companies adhere to a set of principles (called the *generally accepted accounting principles*) that define how to fairly portray a company's economic condition. The task of interest here – second

ULI LIBRARY, ORLANDO, FL 32816

partner review – is a quality control procedure performed by a partner in an accounting firm at the end of the audit of a client. The second partner does not direct the audit. The objective of the review is to validate the conclusion (called an audit *opinion*) that has been reached by the partner in charge of the client audit.

Several different opinions are possible. An *unqualified* opinion means that the auditor believes that the financial statements issued by a client company are fair. An *adverse* opinion means that the auditor believes that there are substantial departures from a fair portrayal of the economic conditions of the company. Because the public and other users of the financial statements largely rely on the auditors' professional judgment, it is paramount for management to obtain an unqualified opinion.

When management successfully manipulates a company's financial statements, both the public and the auditors stand to lose. The users of the financial statements lose financial resources because decisions are made on the basis of deceptive information. When (and if) the manipulation is uncovered, the auditor loses professional reputation and may become the target of the stakeholders' lawsuits (Albrecht, Wernz, & Williams, 1995). To avoid these losses, the auditor needs to manage the risk that the financial officers of the company have manipulated financial information. Activities designed to detect manipulations of information, such as the second partner review, are a means to manage this risk.

For the auditor, thinking about the thinking of other agents involved in the communication of financial information is a determinant of detection success (Grazioli, 1997; Johnson et al., 1993). To do this, the auditor needs to understand management's possible goals and the possible manipulations that can be used to achieve them.

Why Is Managing Risk in Social Exchange Difficult?

Agents in social exchange, including traders and auditors, are faced with the problem of managing the potential for loss posed by interaction with others. Solving this ill-defined problem is difficult. If it were easy to manage risk in social exchange, confidence men (and lawyers) would be out of work.

Shanteau's (1992) theory of expert competence suggests why this problem is difficult to solve, even in settings where problem solvers enjoy high motivation, resources, and training. Shanteau enumerates several task characteristics that are expected to make a task intrinsically

difficult and that are associated with correspondingly poor levels of "expert" performance. Six of these characteristics are relevant to the management of risk in social exchange and would be expected to lead to poor performance in the tasks we address:

- dynamic (changeable) stimuli
- unique tasks
- decisions about behavior
- experts' disagreement on stimuli
- subjective analysis only
- unavailability of feedback

Dynamic stimuli and unique (low base-rate) tasks are sources of novelty. Novelty makes task performance difficult because the task cannot easily be seen as an instance of a familiar class of problems for which the solution is known. Decisions about behavior, disagreements on stimuli, and subjective analyses all lack objective standards. A lack of feedback or objective standards for comparison makes task performance problematic. Without feedback or standards, it is often difficult to define the solution or to judge its proximity.

Most of the six characteristics identified by Shanteau are likely to be present to some degree in trading and auditing settings. Dynamic stimuli in unique configurations, disagreements about stimuli, and decision making about behavior are salient in the spot markets. Decisions about behavior, disagreements about stimuli, lack of feedback, and subjective analysis are salient in detecting manipulations of financial information.

These considerations lead us to expect that few, if any, experts would emerge in the spot markets or in the detection of financial manipulations. Rather, we would expect the level of performance of experts to be relatively poor. This is, in fact, what we observe in the two settings. Available data on the currency markets (Smith, 1996) and fraud detection (Albrecht et al., 1995) show that it is difficult to beat the market systematically or to catch con artists.

By contrast, the same studies have also shown that there is a small number of individuals who seem to be relatively skillful at managing risk in social exchanges (Grazioli, 1997; Johnson, Grazioli, Jamal, & Zualkernan, 1992; Smith, 1996). In the next sections, we propose a theory and a model of the information processing behavior of those individuals who perform competently at the task of managing risk in social exchange.

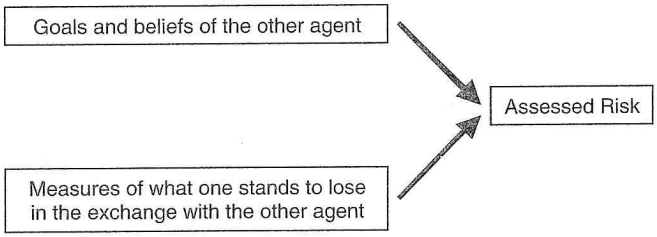


Figure 3.1 Risk assessment in social exchange.

Theory of Assessing Risk in Social Exchange

Drawing from early work by Einhorn and Hogarth (1986), we propose that an individual's assessment of risk in social exchange is influenced by two factors: an understanding of the other agent and a representation of what the individual stands to lose. The individual's representation of what might be lost in the exchange provides the scale for the process of risk assessment. The scale of risk is set high when the individual has a lot to lose. Understanding the other agent's goals and beliefs provides the information to be compared. Figure 3.1 illustrates the main components of risk assessment in social exchange.

We assume that risk assessment is done by taking what Daniel Dennett has called the *intentional stance*:

> Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many – but not all – instances yield a decision about what the agent ought to do; that is what you predict the agent will do. (Dennett, 1987, p. 17)

The strategy takes two steps. First, the agent assigns to a *system* the goals and beliefs that the system "ought to have," given its environment and assuming that it is rational. The goals that a system ought to have are those that directly or instrumentally serve to further its needs and wants. Similarly, its beliefs are assumed to be those it ought to have about its environment: beliefs that match an objective observer's definition of truth.

In the second step, the agent predicts the system's behavior in the light of the assigned beliefs and goals, given the assumption that the system is

rational (Newell, 1982). Taking the intentional stance works "because we are close enough approximations of optimal cognitive design" (Dennett, 1987). Its power consists in the reduction of the set of behaviors that it makes sense to expect.

In social environments, the system may be another agent (e.g., an audit client) or an aggregate agent (e.g., the market). Dennett's analysis suggests that agents involved in social exchange will treat the beliefs and goals of others as causes of their behavior. Attributing beliefs and goals to others vastly simplifies the task of understanding (predicting) their (future) behavior. Understanding and predicting the behavior of others enables the agent to invoke domain knowledge to anticipate losses and take actions that can counter (the effect of) actions taken by others.

Although Dennett is persuasive in arguing that taking the intentional stance is a powerful strategy for understanding and predicting the behaviors of others, he provides only limited detail about the processing that supports it. In what follows, we shall suppose that the process underlying the intentional stance is that of *ascription*. We define ascription as the generation of an inference of the goals and beliefs of an agent, given knowledge about this agent, cues to its behavior, and the environment in which it operates. Ascribing the behavior of others enables an agent to take action to manage the risk the others pose.

As a limitation, we recognize that the process of ascription described here is one of many decision processes required to manage risk in social exchange. Processing of task-specific domain knowledge is the usual focus of studies of expert performance (e.g., Ericsson & Smith, 1991; Peters, 1990). Our focus is on the process of ascription itself.

In the following section, we consider a general model of the ascription process derived from work in social psychology and cognition. We then proceed to test domain-specific models of competent decision making in social exchange that include ascription mechanisms.

Dispositional Attribution

Our model of ascription borrows from the work of Trope (1986; Trope, Cohen, & Maoz, 1988) and other social psychologists who wrote on attribution theory. Attribution theory deals with how the social perceiver uses information to arrive at a causal explanation for events. The theory examines how information is gathered and how this information is combined to form a causal judgment (Fiske & Taylor, 1991).

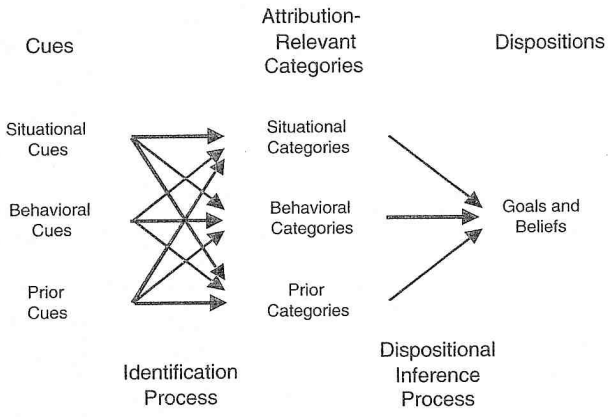


Figure 3.2 Ascription. (Adapted from Trope, 1986.)

Several models have been proposed to account for the attribution process (e.g., Heider, 1958; Jones & Harris, 1967; Kelley, 1971; Kruglanski, 1980). Reckers, Wong-On-Wing, and Krull (1992) have applied the theory of correspondent inferences (a variant of attribution theory) to analyze auditors' assessment of management's dispositions.

Trope has proposed a domain-independent two-stage model of dispositional (intentional) attribution. Because of its generality, we adopt Trope's model as a framework to describe the basic processes and representations that compose ascription in social exchange. Whereas for Trope a *disposition* may be an attitude, a motive, a personality trait, or an ability, we restrict the term to intentional states, specifically to the other agent's goals and beliefs. We also restrict our use of the term *Agent* to the individual doing the ascription. We use the term *Other* to refer to the target of the Agent's ascription process.

Figure 3.2 presents Trope's (1986) two-stage model of dispositional attribution. From left to right, the elements of the model are (1) the inputs to the ascription process – cues from the environment to the Other's dispositions, (2) the first stage of processing, called *identification*, (3) attribution-relevant categories, (4) the second stage of processing, called *dispositional inference*, and (5) the output of the ascription process – the Agent's attribution of the Other's goals and beliefs.

Trope distinguishes three types of cues: situational, behavioral, and prior. Situational cues are cues from the environment, such as economic forces or industry characteristics, that are assumed to be beyond the control of both the Agent and the Other and to which both are likely to respond. Behavioral cues are actions or events that are assumed to be

controlled or influenced by the Other. Prior cues are past information about the Other.

Three parallel classes of constructs mediate the two stages of processing. These constructs are domain-relevant categories that the Agent believes either reveal or influence the Other's goals and beliefs. Together, these categories comprise the Agent's cognitive representation of how cues in the environment either drive or are driven by the Other's disposition. The categories provide meaning for the cues and provide contextual information for the subsequent identification of other incoming cues.

Two stages of processing link the cues to the categories and the categories to the dispositions. The first stage is *spontaneous* identification that associates cues with attribution-relevant categories. The process of identification maps cues into relevant categories (e.g., classifying a behavior as aggressive or secretive). According to Trope, identification is *instantaneous* and context driven:

> [G]iven that the identification process is highly practiced and closely related to the properties of the stimulus information, it is likely to be performed unconsciously and automatically.... (Trope, 1986, p. 241; see also Cosmides & Tooby, 1995)

The second stage of processing is *deliberate* dispositional inference that integrates the identified categories to assign a disposition to the Other (e.g., the intent to be an aggressive seller or to be deceitful). The assigned disposition is a domain-relevant goal or belief that the Other ought to have in light of the current situation.

Dispositional inference assesses the cumulative impact of the attribution-relevant categories. In Trope's words, inference is deliberate and driven by *causal schemata* that specify the procedures for arriving at an attribution of disposition to the Other. As Trope and others illustrate, the causal schemata are often relatively simple algorithmic expressions (Cheng & Holyoak, 1985; Cosmides, 1985). We suggest that appropriate inference procedures in social exchange are fully specified by each of the attribution-relevant categories, making their invocation relatively straightforward.

The original work by Trope is the analytical tool that we use to sharpen our understanding of how goals and beliefs are ascribed to Others in specific domains of human activity. We now apply Trope's framework first to spot currency trading and then to second partner review.

UNIVERSITY OF ICELAND, FLÓRA

Ascription in Spot Currency Trading

Smith (1996) developed and tested an information processing model of risk management in spot currency trading in a series of experiments with professional traders as subjects. The first experiment informed the specification of a trading model. A computer implementation of the model (1) emulated both profitable and unprofitable trading by the traders in the first experiment and (2) predicted their behavior in the second experiment. Smith concluded from this work that the knowledge that supports traders' ability to anticipate the market is a component of trading competence. Here we employ an example from one trader's concurrent verbal report to illustrate how anticipating the market is an exemplar of Trope's general model of ascription.

Experimentation. The illustration requires a brief preface on the experiment that elicited it. Three professional spot currency traders from a mid-sized Twin Cities commercial bank agreed to participate in the experiment. The traders' task was to do their job: to monitor positions in two currencies, to make trades for their bank's customers, to generate profits for their bank, and to square up by the end of the trading day (e.g., 24:00 Greenwich Mean Time, GMT). During the experiment, the traders' thinking processes were captured by asking them to think aloud and by recording their verbalizations. The experiments were conducted in 1993 and 1994, before the introduction of the euro.

The experimental materials were eight scripted spot-market scenarios that compressed realistic world events into compelling trading days. The scenarios were scripted to challenge the traders and to elicit decision making that could be used to generate hypotheses about the organization of the traders' information processing. The scenarios had two parts, an *overnight report* and a simulated *trading day*. The overnight report portrayed the current global socioeconomic and political conditions that provide the context for the trading day. The trader began each session by reading a paper copy of the overnight report. During the trading day, prices in two currencies were continually updated on a price screen, headlines periodically scrolled across a news screen, and the "telephone" repeatedly rang with requests to quote prices. The trader was able at any time to toggle among screens and to initiate a trade.

A commercial PC-based foreign exchange training package (FXPlus, Chisholm-Roth, Ltd.) made it possible to script and present the trading

days to participants. The training package simulates the Reuters trading desk. It has five alternate display screens: three information displays – prices, news, and position – and two trading screens – quoting and dealing. The trader toggled among the five screens to obtain different types of information and to make trades.

Example. The overnight report for the scenario excerpted here indicates that the U.S. Department of Commerce would release data on wholesale prices sometime during the day and that the market expected prices to rise by 0.4%. The excerpt begins during the trading day with the trader reading the headline about U.S. wholesale prices.

Session Time (GMT)

News Screen

6:55 14:10 Washington, 17 May – US wholesale prices rose 0.6 pct in April, a slight increase over the 0.5 pct rise in March, Department of Commerce said.

Transcript

6:58 "Wholesale prices up point six as expected. Wait, wholesale [scans overnight report] zero point four, so they're not going to like the dollar on that one

7:02 and I'm okay against Swiss; Let's get rid of some yen"

In Trope's terminology, this excerpt shows the trader using a situational cue (see Figure 3.2) and a prior cue to invoke a situational category and to ascribe a belief to the market: "they're not going to like the dollar on that one." The headline is the situational cue. The value the market was said to expect is the prior cue. The trader read the headline and interpreted it by comparing the value reported in the headline (0.6%) to the value the market was said to expect (0.4%). The standard for comparison is the situational category. The category for news about inflation indicates that an increase in inflation (e.g., actual minus prior, 0.6% – 0.4%) is bearish for the relevant currency. This standard for comparison prompted the trader to attribute a belief to the market given the news: "they're not going to like the dollar on that one."

The trader's decision making did not stop with the ascription. He moved on to use the ascribed belief to infer the direction in which the market would drive prices (down) and to assess the risk of holding his positions. The trader concluded that the anticipated selling would help his short U.S. dollar–Swiss franc position but hurt his long dollar–yen position, which provoked his decision to trade the dollar–yen. The

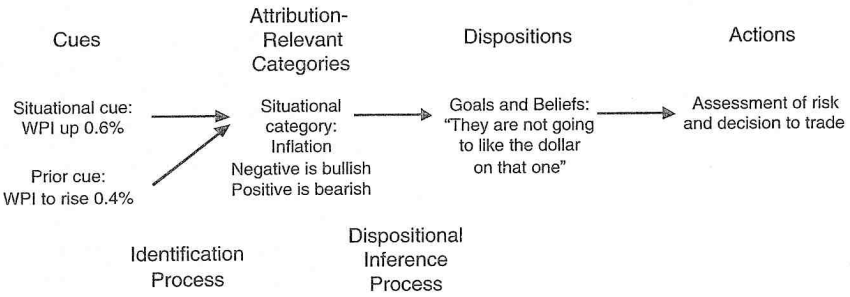


Figure 3.3 The instance of ascription seen in the trader's concurrent verbal report.

ascription of a disposition to the market – "they're not going to like the dollar on that one" – is the key step in the process of deciding how to manage the risk of holding these particular positions in the next few minutes.

Traders ascribe beliefs to the market in order to decide how to manage risk. Figure 3.3 places the instance of ascription seen in the excerpt within the context of our general model (Figure 3.2). The discussion that follows unpacks the components of the ascription process observed in spot currency trading.

Cues. Situational cues are headlines, rumors, and news stories. The spot markets live and die by fact and rumor. To meet the markets' thirst for news, the wire services relay headlines from around the world as soon as they are made. Some headlines appear at scheduled times. Foremost among scheduled news items are government releases of economic indicators like the wholesale price index (WPI) in the previous example.

Market activity tends to slow prior to the scheduled release of a key index or announcement. The Associated Press can be trusted to supply justification for a slowdown in the markets. For example:

> Many dealers with mark holdings have moved to the sidelines ahead of Thursday's meeting of Germany's central bank. Members of the Bundesbank are not expected to take any specific action at the meeting, but dealers are concerned the German interest rates could eventually come down. (June 29, 1993)

Other headlines, like the dissolution of the Japanese Diet or the start of Operation Desert Storm, follow no schedule. Trading activity can be

driven to a frenzy by unexpected events like the following:

> Tuesday 18:00, London – BRITAIN WITHDRAWS FROM EMS
> Prime Minister Major has announced that Britain has dropped out, at least temporarily, of the European monetary system. He added "massive speculative flow has continued to disrupt the functioning of the exchange-rate mechanism."

Often the market does not bother to distinguish between news and rumor (false reports) about a topic (Kurtzman, 1993). The chief difference is that the market tends to hold its belief a little longer when the report is true.

Behavioral cues are the direction of prices. Prices are driven by the consensus of market opinion. Price formation is the behavior traders work to understand. In the example, the trader did not use a behavioral cue.

Prior cues are the trader's preexisting expectations for the direction of prices. In the example, the prior was provided to the trader in the form of an overnight report. Banks prepare overnight reports every night to help their traders to anticipate the market. Traders read the report at the beginning of their trading day and before they begin to trade. The report, condensed from wire service and internal analyses, attempts to define the day's market context. The report lists scheduled headlines for the day and week ahead, consensus predictions about scheduled headlines, and summaries of market activity during the previous day and week.

Consensus predictions about scheduled headlines can seriously skew a market's opinion. When actual events turn out to differ from consensus prediction, the market response can be draconian. In the previous excerpt, the actual event, the situational cue (WPI up 0.6%), differed from the prior cue (WPI to rise 0.4%). The disparity led the trader to expect a sharp market response.

Situational, behavioral, and prior cues are the initial input to the ascription process depicted in Figure 3.2. Ascription is decomposed into two related subprocesses. First, the cues are used to generate attribution-relevant categories (identification). In turn, categories are used to generate dispositions (dispositional inference). The next subsections discuss the components of Trope's model in the context of currency trading.

The Identification Process. Identification is the first stage of processing in Trope's general model for ascription. At the trading desk, identification

UCF LIBRARY, ORLANDO, FL 32816

defines membership of an instance in a class. In the previous example, the trader associates the headline about the U.S. WPI with the class of news Inflation (a situational category). Trope describes the identification process as *automatic*. This description, though facile, appears to apply to spot currency traders. None of the traders in Smith's study ever paused to determine what kind of news a particular headline was. The apparent automaticity of the identification process is the product of years of experience at the trading desk.

Attribution-Relevant Categories. Based on his study of currency traders at the trading desk and in laboratory scenarios, Smith proposed that traders invoke 10 categories of news to interpret headlines in order to anticipate the market. The categories, listed in Table 3.1, are intended to be mutually exclusive and exhaustive. They are both a source of generality for the ascription process and a principled account of the traders' knowledge about the impact of world events on the market's behavior.

The first column in Table 3.1 contains the names given to the 10 categories. The second column contains illustrative members of the categories. Each member represents a generic headline topic, that is, the WPI rather than a specific headline (e.g., the U.S. WPI for March). The third column specifies the appropriate algorithm for comparing situational and prior cues that are members of that category. The fourth column lists the values of implicit standards to be invoked in the absence of prior cues.

The first four categories – economic growth, political uncertainty, inflation, and central bank monetary/government fiscal policy – are the most common. Arguably, these are situational categories. News about economic growth and central bank policy are both assessed using the specification for comparison "positive change is good for the relevant currency," that is, the market can be expected to react favorably to news of increasing growth and to news of rising interest rates. The two categories are distinguished by the implicit market standards the traders invoked to anticipate the market. The traders assessed news about economic growth by checking for a positive or negative value; they acted as if the standard for comparison was zero percent. In contrast, the traders assessed news about central bank policy by checking for positive or negative change regardless of value; they acted as if there were no absolute standard for comparison.

The traders consistently inferred that the market would respond quickly and strongly to devalue the currency of a country that broadcast

Table 3.1. *Categories of News Heeded by Currency Traders*

Category	Sample Topics (Members of the Category)	Specification for Comparison	Market Standards (Implicit Standards for Comparison in the Absence of Priors)
Situational			
Economic growth	Gross domestic product	Positive change is bullish for the relevant currency (good when up)	Bullish when above 0%
Political uncertainty	President rushed to hospital	Necessarily bearish for the relevant currency; for priors: buy the rumor, sell the fact	
Inflation	Wholesale price index (WPI)	Negative change is bullish for the relevant currency (good when down)	Bullish when below 2% year-on-year
Central bank monetary policy/ government fiscal policy	Interest rates Foreign trade	Positive change is bullish for the relevant currency (good when up)	
Unemployment	Nonfarm payroll data	Negative change is bullish for the relevant currency (good when down)	Bullish when below 6% year-on-year
Immaterial	Yeltsin to visit Minnesota	Ignored	
Prior			
Old news	Central bank bought paper	Ignored	
Forecasts	Estimate of next quarter's WPI	Ignored	
Behavioral			
Up trend	Currency strong Peace accord	Necessarily bullish for the relevant currency	
Down trend	Currency weak	Necessarily bearish for the relevant currency	

UCF LIBRARY, ORLANDO, FL 32816

news of political uncertainty. The basis for their decision was not so much a true comparison as an outright classification of political uncertainty as bad news. There is no market standard for news about political uncertainty. The diversity of potential topics – from beheadings to oil embargos – precludes a single market standard.

The third category of news is inflation. Topics in this category are dominated by indices released monthly by government agencies, such as the WPI. Headlines about these topics invariably contain a quantitative value, such as "The U.S. consumer price index rose by 0.6% in March." The traders responded to headlines about inflation by (1) comparing the value announced in the headline to a standard for comparison and by (2) inferring that the market will believe the headline is beneficial for the relevant currency if and only if the announced value is less than or equal to the standard. When a prior cue told them the value the market was said to expect, the expected value acted as the standard for comparison. For example:

Session Time (GMT)

14:00 Washington, 18 May – The U.S. consumer price index rose by 0.6 pct in March, up by 9.7 pct year on year, Commerce Department reported.

15:48 "The US CPI is up point 6. That is as expected."

The trader proceeded to take no action. His (lack of) behavior suggests that he invoked knowledge that the market does not punish the relevant currency when its expectations are met.

The traders responded to topics in the fifth category of news – unemployment – in much the same was as they responded to news about inflation. They (1) compared the gist of the headline to an expected value (if they had one) or to an implicit market standard and (2) inferred that the market would believe the headline was beneficial for the relevant currency if and only if the announced value was less than the standard. Traders use prior cues to infer what the market expects, and in turn use the ascribed expectations as standards for comparison with subsequent information.

The immaterial category contains headlines about political events that drew no response from the traders. Membership in the category is unexpectedly large. Roughly half of the headlines were written to investigate the traders' response to news of political uncertainty. The common theme uniting these headlines is bluster, diatribe, or innuendo unlikely to lead to action. A student of political science might be able to use the

distinctions the traders drew between material and immaterial political topics to inform a model of the materiality of news. This research goes no further than to suggest that the traders treated the market as a keen judge of the support behind the rhetoric.

The old news category contains headlines that revise historic (prior) data about topics in the categories economic growth, inflation, central bank policy, and unemployment. Although traders read headlines that contain revisions to data, they give no indication that they use them to anticipate the market. Similarly, traders appear to ignore forecasts, headlines that project future values for news about economic growth, inflation, central bank policy, and unemployment. The traders' specification for comparison was, in effect, to ignore old news and forecasts.

The last two categories – up trend and down trend – are used to assess behavioral cues. Up trend provides the specification for dispositional inference based on rising prices. Down trend provides the specification for dispositional inference based on falling prices.

The Dispositional Inference Process. The specifications for comparison shown in Table 3.1 indicate the appropriate algorithm for making an inference about the belief the market will form given a situational cue (headline) with or without an associated prior cue (expectation). The knowledge contained in the 10 categories of news reduces the process of dispositional inference to a straightforward comparison of evidence. This knowledge supports the comparison of situational and prior cues (and the assessment of behavioral cues) and is one of the foundations of competent decision making in spot currency trading.

Ascription in Second Partner Review

Johnson, Grazioli, Jamal, and Berryman (2001) have recently argued that understanding how Others think and act is a key for success in detecting financial manipulations. Their work has applied Dennett's intentional stance strategy to the field of auditing (1992, 1993) and includes the development and testing of several information processing models of second partner review.

The 1992 and 2001 studies have examined the responses of auditors engaged in the second partner review of a set of financial statements. Unknown to the auditors, some of the reviewed statements were fraudulent. The auditors were asked to "think aloud" as they reviewed the

UCF LIBRARY, ORLANDO, FL 32816

statements. Their verbal protocols were tape-recorded, transcribed, and coded with reliable coding techniques (Johnson et al., 1992).

The coded protocols were the basis for developing a process model (a computer program) of competent detection of financial manipulations in second partner review. At its core, the model embodies two main skills: (1) the ability to interpret cues in the financial statement in light of what management might want to accomplish and (2) the ability to combine these cues to form a final opinion. The 1992 study showed that experimentally enabling or disabling the expression of these two skills in the model allowed the model to generate a range of behavior that is similar to the range observed in the auditors who participated in the experiments.

In particular, it was found that the models (and auditors) that ascribe goals to the management of a client company are more successful in detecting financial manipulations than the models that do not. As was done for the currency traders, we illustrate this point by commenting on one of the auditor's concurrent verbal report and by using Trope's general model as a frame of reference.

Experimentation. The data presented here are derived from the 2001 study. Twenty-four senior partners from the major accounting firms were asked to perform a second partner review of five sets of financial statements. All five sets described companies for which the partner in charge of the audit intended to issue an unqualified (i.e., "clean") opinion.

All the statements described real companies. Four of the five statements were fraudulently manipulated by the company's management for the purpose of making the company appear more profitable than it actually was. The following example is excerpted from the protocol of subject S17.

Example. S17 examined the financial statements of "Surgical Products," a fictitious name for a company operating in the medical products industry. The company severely misstated its earnings by means of several manipulations, including overstating its inventory, opportunistically changing the accounting treatment of its research and development expenses, and recording shipments of products to wholesalers and customers as sales (Johnson et al., 1992, 2001).

S17 began his analysis by reading the general description of the company's products and markets as provided by management.



Figure 3.4 The instance of ascription seen in the auditor's concurrent verbal report.

S17 ... Okay, they've got a line of products that has one very good characteristic to them from the standpoint of a successful company and that is that they are disposable, so they sell and the user only gets one use out of them.
... Sales of their "autosuture" products which is one line [...] account for all, substantively all, or 99% of company profits, so those other products are not necessarily doing um that well ... either that or that they are brand new products and they haven't [...] achieved their market share yet, but that's something to look at in terms of realizability of whatever investment they have in them.
... Okay, attributes its growth in profits this year to three elements. [...] Ah – one is selling price increases of 10% which is pretty strong increase and then growth in volume and ah new products....
[reading about the competition] ... It is basically a small player in this particular market for wound closure [...] which makes it surprising that they have been able to sustain a 10% price increase but ah it could be that their products are unique in the marketplace.

The instance of the ascription process presented in the protocol within the context of Trope's process model is graphically summarized in Figure 3.4. In Trope's terms, the company's products, markets, and

LIBRARY, ORLANDO, FL 32816

competition are situational cues that the auditor uses to identify corresponding situational ("small player," "one-product company") and behavioral ("highly innovative, risky") categories.

More generally, *situational cues* are information about the company's internal and external economic environment. For instance, the amount of inventory is a situational cue describing the company's internal environment; the level of market competition is a situational cue describing the company's external environment. Most of the current balances in the income statement and balance sheet are situational cues. The past balances are *prior cues*. The following is an excerpt from S17's analysis of the inventory balances:

S17 ... Large increase in inventories, primarily in the finished goods area ...
cost of goods sold is up 48% ...
inventories are up 76% ...
again that includes quite a bit of raw materials, which means that the grossed invent[ory] ... the grossed up inventories after completion are up substantially greater than 76% [...]
so I'd like to understand what the strategy is in terms of building inventory and why so much finished goods isn't being sold
this goes back to the issue of getting a breakdown by product since, as noted earlier, only 1% of the company's sales comes from a number of different products....
make sure that they do not have any realizable value problems or any excess inventory problem.

S17 exploits the fact that data are organized in terms of current versus prior reporting years to compute percentages of change, which in turn allows him to generate the situational category "inventory is too high," as well as the behavioral categories "they are not selling their product" and "they are not managing inventory well." The chronological structure of the financial statements offers a rich, well-organized set of prior and situational cues from which to infer situational and behavioral categories.

S17 ... Now reading the um accounting policy on molds and dies well they've got um what is either a correction of an error accounted for in the current period or a change in accounting which they are calling ...

which they're calling a development of improved information
it's a . . . basically capitalization of engineering labor
certain raw material costs . . . hum . . . it is really quite significant
since the company's pretax income is essentially flat [. . .] so
you've got a fairly dramatic change in accounting which affects
the consistency of the costs elements in the income statement
without that change the earnings would have apparently been
down significantly in the current year

In this passage, S17 uses a behavioral cue (the changes made to the accounting policy) and a situational cue (the impact on the bottom line) as a basis to derive a behavioral category "aggressive accounting changes that make the company look more profitable." More generally, *behavioral cues* are information on management's actions and reporting choices.

S17 . . . Now reading the Research and Development accounting policy and R and D cost are down dramatically . . . which is odd considering their statements about developing new products and the tremendous investment they are making in molds and dies associated with the new products

The comparison of a prior cue with a situational cue (past and present R&D costs) generates the behavioral category "lowering research spending." However, this category is in stark contradiction with a previously identified behavioral cue, that is, the development of new products. S17 solves the contradiction between cues and categories by combining them with the "aggressive accounting changes" category previously generated:

S17 . . . Could very well be that they have shifted over some of their effort to ah actually get these products into production . . .

which is one the manipulations actually perpetrated by the company's management to increase reported net income. This ascribed goal is consistent with other cues and previously generated dispositions:

S17 The company's got a strong equity position – increased from 54 million to 100 million . . . substantial part of the increase is coming from a public offering
they had public offerings both in 19X1 and 19X2
again suggest they are very sensitive to maintaining good earnings trends and good earnings.

By the end of his review, S17 identified many other items in the financial statements of Surgical Products that he interpreted as manipulated or as evidence that manipulations took place. In his concluding remarks he stated:

S17 ... It is unlikely that we would issue an unqualified report [...] if everything is a problem, then I'd say that they've got a pervasive problem with the entire financial statements and I would say that it should be an adverse opinion

which is the correct way to manage the risk posed by management of Surgical Products.

The Identification Process. Johnson et al. (1992, 1993, 2001) argue that auditors identify ascription-relevant categories by means of a comparison between observed cues (i.e., the available financial information) and expectations about these cues. Some of these expectations are independent of the financial information provided by the company and are the result of the auditor's knowledge of his or her domain. Other expectations are generated as a function of the information presented in the financial statement.

There is both deliberate and automatic information processing in the identification process used by the auditors. S17 began his review of the inventory balance by deliberately calculating a percentage of change between past and current values of the inventory balance. A more automatic comparison of the percentage of change and its expected value led him to categorize the balance as too high.

The Dispositional Inference Process. The cited studies also specify algorithms by which the dispositional inference process may work. Auditors generate hypotheses about the manager's goals and actions in response to the concurrent presence of three cognitive representations: (1) a violated expectation (an anomaly), (2) an assessment that management might have caused the anomaly, and (3) an assessment of the relevance of the anomaly to possible managerial goals.

For example, the observed value of inventory violated S17's expectation of what that value should be. This anomaly prompted the explanation that the difference may have resulted from a mishandling of the accounting procedures that determined the current balance for inventory, and specifically from neglecting "realizable value problems or any excess inventory problem." In addition, the anomaly contributed to the general goal of "maintaining good earnings trends and good earnings."

Jointly, these three elements led to an ascription to management of the specific goal of intentionally overstating inventory. This ascription was a critical component in the final outcome of S17's review of the financial statements presented by Surgical Products.

The experimental data presented thus far have illustrated how professionals who need to manage risk in social exchanges succeed at their tasks by ascribing goals and beliefs to Others. The next section deepens our investigation of the ascription process by experimenting with information processing models of currency traders and auditors.

The Role of Ascription in Managing Social Exchange Risk

The behavioral data elicited by our simulated market scenarios (spot currency trading) and financial reports (second partner review) provided the basis for constructing models of the information processes underlying the performance of experienced traders and auditors. A description of the models allows us to move beyond the anecdotal analyses presented so far and forces us to be specific about the processing that supports ascription. It also sets the stage for the creation of two *ablated* models. The ablated models are the original models with the implementation of ascription removed or modified. As a last step, we assess the critical role of ascription in managing social exchange risk by contrasting the behavior of the full and ablated models with that of our experimental subjects.

A Model of Spot Currency Trading

The development of the model of currency trading borrows the paradigm of process control from human factors psychology and industrial engineering (Broadbent, 1977; Edwards & Lees, 1974; Moray, 1986; Parasuraman, 1986). The goal of the process control Agent is to keep a process within operational limits as its environment continually changes. The Agent makes decisions about whether, when, and how to take action to control a dynamic process (Wiener, 1984). A familiar example is driving a car; the driver's goal is to keep the car on the road and at a safe distance from Others as traffic conditions vary. The driver's decisions are designed to guide the car down the road. The process control Agent's decisions manage the risk engendered by the interaction of the process with its environment.

The process the spot currency trader seeks to control is his or her position – the value of assets exposed in the international markets. As a process control Agent, the trader samples market information and makes decisions about whether, when, and how to make trades that keep that position within limits of acceptable risk.

There are two general algorithms for process control – feedback control and feedforward control (Brehmer, 1992). An Agent who relies on feedback control samples information about the current state of the process and its environment in order to respond to the current risks and take appropriate action. An Agent who relies on feedforward control samples information that affords anticipation of the future state of the process and its environment in order to infer future risks and take appropriate action. In many domains, skilled Agents display a mixture of the two algorithms (e.g., Bainbridge, 1974).

Our model of spot currency trading (Smith, 1996) proposes that traders adopt a mixture of feedback and feedforward control. Their method of feedback control responds to changing prices with an assessment of the risk facing their positions at that instant and with a judgment of whether and when to make a trade. Their method of feedforward control anticipates the market in order to infer the risks their positions are likely to face in the next few minutes. The focus here is exclusively on the traders' method for feedforward control.

Anticipating the Market: Feedforward Control. We argue that traders adopt the intentional stance to anticipate the market. Specifically, we assume that they address the market as if it were a single rational Other that, on average and over the long run, acts in manner that directly or instrumentally increases its profitability. Adopting the intentional stance with respect to a composite Other requires the assumption that individual Others within that composite share the same goal. We assume that all traders in the spot market share the same goal: increased profitability.

Attributing to the market the goal of increasing profits enables a powerful application of the intentional stance. The trader takes the intentional stance to anticipate the beliefs the market will form in order to predict the direction in which it will drive prices. The basis for the market's beliefs and for the trader's ascription of those beliefs is information about global socioeconomic and political events. This information takes the form of situational cues and prior cues. The situational cues are headlines broadcast by wire services to news screens at trading desks around the world. The prior cues are expectations for the information the market will find when it reads headlines.

Anticipating the market's beliefs allows traders to infer how the market will behave given the beliefs it has formed. Attributing beliefs to the market vastly simplifies the traders' task of understanding (predicting) the market's (future) behavior. Understanding and predicting the direction in which the market will drive prices in response to the headlines it reads enables traders to assess the risks the market's behavior may pose to their positions and to make trades that manage those risks.

Performance Model. To test our model of currency trading we developed a rule-based system that trades foreign currencies in six markets (e.g., the U.S. dollar vs. the German mark). The system, called SPOT, responds to the same stimuli presented to the traders in Smith's laboratory experiments (1996). For each of the eight trading scenarios employed in the experiments, SPOT processes sequentially the news in the overnight report, the headlines, the hourly prices for the currencies, and the request for quotes submitted to the human traders. In response to these inputs, SPOT assumes positions in the currency markets in which it operates. Appendix A describes SPOT's algorithms and technical implementation.

Ablated Model. To ascertain the impact on trading behavior of the ability to ascribe beliefs to the market, we created a version of SPOT that does not ascribe beliefs to the market. Specifically, we modified the original program so that it never reads headlines. This ablation makes SPOT blind to situational cues. The inability to respond to situational cues makes it impossible for SPOT to infer the direction in which the market will drive prices and to assess the future risk to its positions. As a consequence, the ablated model operates in a pure feedback-driven mode. Its decisions to make a trade are prompted by comparing its profits (losses) and the rate of change of prices to static thresholds of acceptable risk. The only measure of risk available to the ablated model is the size of its positions. Accordingly, it invariably directs its attention to the larger position rather than the more newsworthy position.

Results. Table 3.2 presents the profits (losses) generated by the full and ablated models and the professional currency traders in the laboratory experiment. The eight large boxes correspond to the eight trading scenarios. The ninth scenario was used as the subjects' warmup task and was dropped from the analysis.

UCF LIBRARY, ORLANDO, FL 32816

Table 3.2. *Profits (Losses) Incurred by the Traders and the Models*

	German Marks	Italian Lira	British Pounds	French Francs	Japanese Yen	Swiss Francs
Trader 1	$8,171	($2,237)	($4,200)	$121	($53,639)	($4)
Full	$4,966	$148	($7,618)	($2,566)	($26,042)	($3,456)
Ablated	$1,291	($3,447)	($20,750)	($475)	($27,801)	($5,186)
Trader 2	$3,206	($4,961)	$26,678	$4,982	($26,649)	($5,414)
Full	($724)	($3,599)	$15,805	$7,166	($15,476)	($7,664)
Ablated	$1,653	($3,257)	($2,799)	$6,568	($20,966)	($12,574)
Trader 3	($3,086)	($1,102)	$29,606	$3,062	($53,613)	$1,209
Full	($1,874)	($3,029)	$55,323	$4,770	($27,332)	$4,600
Ablated	$1,392	($4,478)	$11,734	$5,139	($35,702)	$4,734
Trader 1	$13,250	$28,748			$27,089	$30,050
Full	$11,574	$20,447			$15,145	$31,680
Ablated	$6,495	$28,217			$6,605	$21,155
Trader 2	$31,086	$33,137			($2,275)	$9,067
Full	$33,470	$16,856			($922)	$15,905
Ablated	$25,363	$4,285			($8,969)	$7,257
Trader 3	$3,877	($6,182)			($16,245)	$686
Full	$7,702	($10,565)			$6,753	$14,936
Ablated	$3,384	($15,543)			($13,139)	$13,283
Trader 1	$13,295	$15,141	($12,775)	$36	($96,990)	($10,915)
Full	$12,602	($5,160)	($5,871)	$2,634	($50,363)	($22,439)
Ablated	($22,076)	$1,991	($12,808)	($5,904)	($81,980)	($13,286)
Trader 2	$17,566	$469			($123,006)	($19,314)
Full	$48,101	($2,423)			($79,963)	($22,007)
Ablated	$14,973	($20,812)			($73,152)	($28,870)
Trader 3	$18,104	($8,359)	($42,190)	$2,464	($40,471)	($10,648)
Full	$12,668	($8,280)	($39,999)	$2,831	($51,049)	($14,835)
Ablated	($4,430)	($8,280)	($29,401)	$2,884	($45,893)	($15,524)

Table 3.3. *Results of the Comparison between Models and Traders*

Result	Interpretation	*n*	%
Ascription improves predictive accuracy and increases profits	Ascription explains profitable trading by the trader	19	42%
Ascription improves predictive accuracy but lessens profits	Ascription explains relatively unprofitable trading behavior	7	16%
Ascription hurts predictive accuracy but increases profits	Modeling suggests that the trader fails to invoke ascription	11	24%
Ascription hurts predictive accuracy and lessens profits	Ascription does not explain trader performance	8	18%

In each scenario, the traders and the models were responsible for managing positions in the two currencies indicated at the top of each column (e.g., marks and lira, yen and Swiss francs). The box for each scenario is divided into three sections, one for each professional trader. The top row in each division represents the profits (losses) generated by the trader for that scenario. For instance, in scenario 1, Trader 1 made a $8,171 profit in the dollar–mark and took a $2,237 loss in the dollar–lira; Trader 3 lost $3,086 in the dollar–mark and $1,102 in the dollar–lira.

Table 3.3 summarizes the results presented in Table 3.2 by presenting counts of (1) predictive accuracy and (2) effectiveness of the models. Predictive accuracy is assessed by measuring how well the models approximate the profits (losses) realized by the traders at the end of the scenario. Effectiveness compares the profits realized by the two models.

The data in Table 3.3 show that the model that includes the ascription mechanism realizes net profits (losses) that are closer in absolute value to the profits (losses) realized by the traders than the ablated model in 26 out of 45 scenarios (58%).[1] In 30 out of 45 scenarios (66%), the model that includes ascription realizes higher profits than the model that does not.

[1] For each of the eight scenarios, there are six comparisons (6 = 2 currencies × 3 traders). Three comparisons are missing from Table 3.3: two for Trader 2 in the lower pound–franc scenario (missing data) and one for Trader 3 in the lira where there was no difference between the two models.

ROLLINS COLLEGE LIBRARY, ORLANDO, FL 32804

The results in Table 3.3 prompt three conclusions. The first conclusion is that ascription generally helps to explain profitable trading behavior. The most frequent finding (42% of the cases – see Table 3.3) is that the full model is better than the ablated model both in terms of profits and in terms of its ability to approximate the outcomes of the experimental subjects.

The second conclusion is that sometimes the traders either abstain from ascribing beliefs to the market or override the implications of ascription. In our data, this happened in about 40% of the cases (24% plus 18%, as shown in the last two rows of Table 3.3). Interestingly, the impact of this behavior on overall profits and losses is not clear: In 11 cases out of 19 (=11 + 8), the traders who did not ascribe improved their profits, yet they reduced them in the remaining 8 cases.

The third conclusion is that ascription is not always an optimal behavior: It appears that in about one-sixth of the cases (16%, as indicated in the second row in Table 3.3), attempting to ascribe beliefs to the markets is not the course of action that maximizes profits.

In sum, the role of ascription in decision making under risk that emerges from these data is significant, yet more complex than indicated by our initial expectations. To sharpen our understanding, the next pair of models do not assume that the subjects either always ascribe or never ascribe. The model of second partner review introduced in the next section treats the use of ascription by the experimental subjects as an exogenous (e.g., independent) variable. In other words, it ascribes goals and actions to management when the subjects do, and does not when they do not. This model will then be compared with an alternative model of second partner review that always applies ascription.

A Model of Detecting Financial Manipulations

The strategy employed to develop models of detecting financial manipulations in second partner review consists of four steps (Johnson et al., 2001). First, we developed a model of successful detection of financial manipulations, which we call the *competence* model. As a second step, we compared the behavior of the model (a computer program) with the behavior of the auditors in our experiments. The comparison revealed a small number of key deviations from the model of success. We call these deviations *errors* (the list of the errors is provided later).

The error data were the basis for formulating hypotheses on the nature of the (flawed) knowledge that leads to error. As the third step

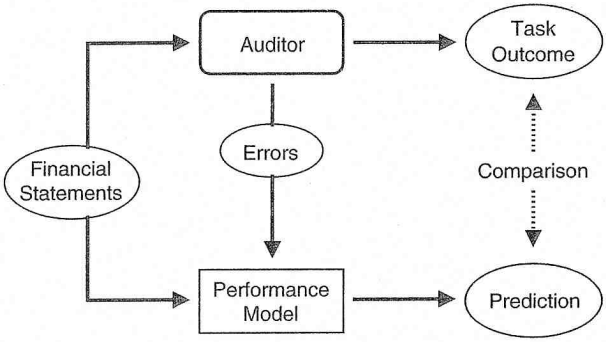


Figure 3.5 Predicting outcomes from observed errors.

in our methodology, we modeled this knowledge as a small number of localized modifications to the competence model. We called these modifications *knowledge bugs* (Van Lehn, 1990) and the model that includes them the *performance model* (Johnson et al., 2001). Individually, each knowledge bug is not sufficient to make the performance model fail the second partner review task. The model (and the subjects) can succeed at the overall task despite making a few errors. The knowledge bugs interact with the characteristics of the task and with each other to determine the final outcome.

Accordingly, the fourth step consists of testing the performance model by using it to predict the outcomes rendered by a new sample of auditors on a set of experimental tasks (Johnson et al., 2001; see also Grazioli, 1997). Given a task and the set of errors made by an individual subject, the model predicts the outcome on the task for that subject. Figure 3.5 shows the technique employed to compute the predictions in step four.

To evaluate the role of ascription in second partner review, we add a fifth analytical step. We omit from the performance model the knowledge bug that causes ascription errors. The resulting *ablated model* takes into account all sources of subject variance from the model of success – that is, all the errors – *except* the variance in ascription behavior (i.e., the ascription errors – more detail is provided later). We expect that this technique will lead to sharper results than the study of currency trading because the performance model of second partner review is a closer approximation to the subjects' ascription behavior. The two alternative versions of the model of currency trading either always ascribe or never ascribe. The performance model of second partner review ascribes goals and actions if and only if the subjects do.

The following sections describe the auditing models, how they ascribe to management the intent to manipulate financial information, and the results of the comparison between the full and ablated versions of the model.

Model of Competent Detection. The development of our model of second partner review draws from theories of deception and its detection (Ceci, Leichtman, & Putnick, 1992; Ekman, 1991; Johnson et al., 1993; Mitchell & Thompson, 1986). Broadly defined, deception is the phenomenon that occurs when an Other – the deceiver – manipulates the environment of an intended victim so that the victim misrepresents the environment and, as a consequence, acts in ways that are more favorable to the deceiver. In second partner review, we look to management as the deceiver and to the auditor as the intended victim of deception. Although the ultimate victim is the public – the users of the manipulated financial statements – management needs the auditor to render a favorable audit opinion.

Deception is detected by realizing that the environment has been manipulated by the deceiver. Our research has proposed that deception is detected by adopting a variation of Dennett's intentional stance. In short, the stance consists of interpreting anomalies in an environment characterized by the presence of a motivated adversary (the potential deceiver) as resulting from the potential deceiver's malicious action.

We have argued that, despite the apparently endless variety of cons and ruses, there is only a small number of general ways to deceive (Bowyer, 1982), which we call *deception tactics*, and that deception is detected by applying a corresponding set of *detection tactics* (Johnson et al., 1993, 2001). Detection tactics are heuristics (cognitive processes and representations) that generate the hypothesis that a specific deceptive manipulation has been attempted by an adversary. Detection tactics are forms of dispositional attribution.

For instance, one general form of deception is *repackaging*, which consists of changing the name or label of an entity (the deception *core*) so that the victim takes it for something else and, as a result, behaves in ways that are more favorable to the deceiver. In the previous example, the management of Surgical Products intended to obtain the auditor's approval of the capitalization (a favorable accounting treatment) of certain expenses by calling these expenses "engineering costs for molds and dies," as opposed to "R&D costs" which have a less favorable accounting treatment.

A list of the detection tactics and further examples in the domain of accounting can be found in Johnson et al. (1993, 2001) and Grazioli (1997).[2] The tactics share features with Trope's model (Figure 3.2). Each tactic is initiated by the observation of an anomaly in the environment and consists of a process that operates on three representations (*categories* in Trope's model): (1) an anomaly is present, (2) the anomaly is functional to the goals of the potential deceiver, and (3) the potential deceiver is capable of generating the anomaly. When all three attribution-relevant categories are present, the intent to deceive (a dispositional attribution) is ascribed to the adversary.

The protocol excerpts presented earlier exemplify the use of the antirepackaging detection tactic. S17 began by noticing that the capitalized molds and dies costs were unusually large, especially in terms of their impact on Surgical's bottom line. In addition, the reduced R&D expense that resulted from the capitalization was functional to one of the ascribed goals of management: showing increasing earnings to the financial markets (Figure 3.4). Finally, S17 commented on the discretionary power of management in labeling those activities one way or another.

The detection tactics have been embedded in a model of successful detection of financial manipulations in second partner review (Johnson et al., 1992, 1993, 2001). The model provides algorithms for the identification of the anomalies that initiate the tactics, algorithms for generating interpretive hypotheses for the anomalies, and algorithms for combining available interpretations of the data.

To test the model, we implemented it as a rule-based system. The system analyzes the same financial statements that were given to the second partner reviewers who participated in our experiments, applies the detection tactics to generate hypotheses about possible manipulations in the statements, evaluates the generated hypotheses, and combines them in an outcome that expresses the extent to which the financial statements are a fair representation of the company's economic condition. Appendix B describes the algorithms of the model and how they were implemented in the rule-based system.

The results of the model's testing have led us to conclude that the competence model hones our understanding of the behavior of the auditors that succeeded in our experimental tasks. However, it is not equally useful for explaining the behavior of the many auditors that did not succeed

[2] Grazioli (1997) identifies six deception tactics and six corresponding detection tactics: masking, dazzling, decoying, mimicking, double-play, and repackaging.

in our experiments (Johnson et al., 1992, 2001). To understand the behavior of all subjects, including those who did not detect the manipulated financial information, we developed a model of human performance in second partner review that includes human error.

Performance Model. The comparison of the process traces of the experimental subjects with the process trace of the competence model described in the previous sections identified four types of errors: (A) failing to heed critical cues in the financial statements, (B) generating an incorrect hypothesis (accounting mistake or insufficient disclosure rather than the hypothesis that a manipulation has occurred), (C) dropping correct hypotheses, and (D) failing to combine available evidence.

The errors of interest here are the failures to generate hypotheses that explain identified inconsistencies in financial statements by means of intentional manipulation by management (error B) – which is, as discussed earlier, an ascription. We identified two variants of ascription error B. Error B1 consists of explaining a manipulated cue as an accounting mistake. Accounting mistakes are, by definition, unintentional. Generating such an hypothesis means that the subject correctly identified an anomaly in the accounting process but failed to recognize it as a purposeful manipulation made by management. Error B2 consists of explaining a manipulated cue as an insufficiently disclosed issue. *Insufficient disclosure* refers to the presentation of financial information. The hypothesis of insufficient disclosure is neutral with respect to intentionality. In this case too, the subject falls short of ascribing the intent to manipulate information to management.

We assume that errors occur because of *knowledge bugs* – localized imperfections in a cognitive Agent's processes and representations that manifest themselves as local errors and can lead to ultimate task failure (Grazioli, 1997; Johnson et al., 2001; Van Lehn, 1990). Accordingly, four knowledge bugs (A, B, C, D) were proposed to account for each of the four observed types of error. The performance model adds these knowledge bugs to the competence model described earlier. The purpose of the performance model is to better emulate subject behavior by incorporating capacities for error. It should be noted that no individual bug is sufficient to make the model fail across cases. It is the interaction among bugs that can make the model fail.

Knowledge bug B causes hypothesis generation errors B1 and B2. Bug B is conceptualized as the inability to conceive of a way that management could have caused a detected inconsistency. It is operationalized

by inhibiting the performance model's access to specific managerial actions that would explain the inconsistency (Grazioli, 1997). As a result of the modification, the model can still identify manipulated information as an anomaly, but it fails to retrieve a managerial action that could have caused the anomaly. This hampers the model's ability to generate the correct hypothesis that the presented financial information has been manipulated, so that, by default, a weaker *accounting mistake* or *insufficient disclosure* hypothesis is generated.[3]

Ablated Model. To evaluate the role of ascription in second partner review, we created an ablated version of the performance model. The full performance model contains all four bugs and can make all four types of errors that characterize human performance. The ablated model does not contain the ascription bug B (but does contain bugs A, C, and D). Because the ablated model does not contain bug B, it *does not* make ascription errors. Comparison of the ablated and full models reveals how much of the models' effectiveness and predictive accuracy depends specifically on modeling ascription behavior (errors).

Results. The impact of ascription on second partner review is evaluated by comparing the outcomes of experienced auditors with the outcomes of the full and ablated performance models. As before, we use two dimensions of model quality: predictive accuracy and task effectiveness. To measure predictive accuracy, the outcomes of the full and ablated model outcomes were compared with data from experimental subjects engaged in the second partner review of four cases. Task effectiveness was measured by comparing the outcomes of the model with the correct solution for each case.

The subjects used for assessing the models are 24 senior partners from major accounting firms who volunteered for the experiment reported in Johnson et al. (2001). Each subject worked on four manipulated cases describing real companies that perpetrated financial fraud. Because 24 subjects worked on four cases, we have a total of 96 task outcomes. Each subject's response was coded as one of three possible outcomes: "M," assigned when the subject rated the case as misleading; "U+," assigned when the subjects felt that the case needed to be qualified; and "U," assigned when the subject felt that the case was a fair representation of

[3] Which one depends on the specific type of the cue (e.g., a cost balance in the income statement or a footnote to the financial statements).

the company's economic condition. Because all cases were fraudulently manipulated, M is the successful outcome, and U and U+ denote failure to detect.

Johnson et al. (2001) report 45 successes and 51 (53%) task failures. As expected, it was found that auditors may succeed in the task despite making some errors (e.g., missing a cue). In the 96 tasks, the 24 subjects made a total of 209 errors. A comparison of the relative frequencies of each type of error shows that error B (the ascription error) was the most frequent, with 114 occurrences (55%).

Table 3.4 presents the ascription errors and the outcomes selected by the subjects in the experiment. For each case submitted to the subjects, we focus on the two major manipulations. Verbal protocol analysis was employed to reveal the types of errors made by each auditor in processing each manipulation (Johnson et al., 2001). For instance, management of the Pharmaceuticals company manipulated their mailing lists and their inventory balances to increase their reported income. The data in the table show that every subject made at least one ascription error B (see, e.g., S2), and some made as many as eight (S12).

An examination of the outcomes selected by each subject in each case shows that the ascription error seems to be frequently associated with failure to detect. About 89% (32/36) of the subjects who made two ascription errors B also failed the task. About 83% (15/18) of the subjects who did not make ascription errors with respect to the major issues in each case succeeded. Error B on the major issues of a case is, however, not fatal: Some subjects identified cases as manipulated despite making two errors B (e.g., S1 on the Pharmaceuticals case). In a related study with a different subject sample, Grazioli (1998) has shown that the occurrence of ascription error B is significantly correlated with the failure to detect.

The error data for each subject, including the error B data presented in Table 3.4, were used to turn "on" or "off" the corresponding bug in the performance model. When the bug was present, the performance model manifested the corresponding error behavior. When the bug was absent, the model defaulted to the competence model and performed correctly. No single error was sufficient to make the model fail: The outcome of the model (i.e., M, U+, or U) on a case was determined by the interaction of the bugs with the characteristics of the specific case.

Table 3.5 presents the results on predictive accuracy of the full and ablated models. Each model predicted the task outcome for each subject on each case in the experiment. The prediction was compared with the outcome selected by the subject as presented in Table 3.4. The full

Table 3.4. *Task Outcomes and Ascription Errors Made by 24 Subjects on Four Fraud Cases*

	Cases (All Manipulated)											
	Pharmaceuticals			Surgical Products			Big John's			America's Family		
Subject	Mailing lists	Inventory	Outcome	Molds & Dies	R&D	Outcome	Pre-opening costs	Inventory	Outcome	Inventory	Goodwill amortization	Outcome
S1	B	B	M	B	B	M		B	M		B	M
S2*			M			M		B	M			M
S3		B	M			M	B	B	U+		B	M
S4		B	M			M	B	B	U+	B	B	U
S5		B	M	B		U		B	M	B	B	U+
S6		B	M	B	B	U+	B	B	U			M
S7		B	M	B		M		B	M	B	B	U+
S8	B	B	U			M	B	B	U+		B	M
S9	B	B	U	B		U+		B	M		B	U+
S10		B	M	B	B	U+		B	U+	B	B	U+
S11		B	U		B	M	B		U	B	B	U
S12	B	B	U+	B	B	M	B	B	U+	B	B	U
S13		B	M	B	B	U+		B	M			M
S14		B	M	B	B	U+		B	M			M
S15			M			U		B	M	B	B	M
S16	B	B	U+			M		B	M	B	B	U+
S17	B	B	U	B		U+		B	M	B	B	U
S18			U	B		U+			M	B	B	U
S19			M			M	B	B	U+	B		U+
S20		B	M	B		U+		B	M	B		U
S21			M	B		U	B	B	U+	B	B	U
S22		B	M	B	B	U+	B	B	U+			U
S23	B	B	U+	B		U+	B	B	U+	B		U
S24	B	B	U	B		U	B	B	U	B		U

Notes: B = the subject made an ascription error in processing the cue. M = the subject rated the case as misleading. U+ = the subject felt that the case needed to be qualified; U = the subject felt that the case was a fair representation of the company's economic condition. Because all cases are fraudulently manipulated, M is the successful outcome; U and U+ denote failures to detect.

Source: Adapted from Johnson et al. (2001).

UCF LIBRARY, ORLANDO, FL 32816

Table 3.5. *Predictive Accuracy of Alternative Models of the Auditor*

Model Version	Interpretation	Case: American Family	Case: Pharma-ceuticals	Case: Surgical Products	Case: Big John's	Total
Full Performance Model	High percentages mean that the outcome agreement of the model that emulates the auditors' ascription errors and the auditors is high.	83% (20/24)	87% (21/24)	83% (20/24)	96% (23/24)	87% (84/96)
Ablated Model (No Bug B)	Lower percentages mean that the outcome agreement of the model that does not emulate the auditors' ascription errors (always ascribes) is lower.	58% (14/24)	75% (18/24)	46% (11/24)	42% (10/24)	55% (53/96)

performance model averaged 87% of correct predictions, the ablated model 55%. The average net loss of predictive power due to ablation is 32%. We conclude that the ability to ascribe goals to a potential adversary explains much of our subjects' overall performance, inclusive of success and failure.

The results on the effectiveness of the full and ablated models are presented in Table 3.6. Effectiveness is defined as the proportion of successes (i.e., selecting an M outcome for each misleading case) over the total number of tasks. For comparison purposes, the first row of Table 3.5 contains the effectiveness of the 24 auditors in the experiment.

The average success rate of the auditors across cases was 48%. The model with the full set of knowledge bugs succeeded 58% of the time across cases; the ablated model succeeded, as expected, much more often: 83% of the time. Table 3.6 prompts the conclusion that ascription errors have a large and negative impact on effectiveness in detecting manipulated financial information. The elimination of flawed ascription knowledge (bug B) increases the average effectiveness of the performance model by 25% (= 83% − 58% in the last column of Table 3.6).

Table 3.6. *Effectiveness of Experimental Subjects and Alternative Models of the Auditor*

Source	Interpretation	Case: American Family	Case: Pharma-ceuticals	Case: Surgical Products	Case: Big John's	Total
Experimental subjects	Auditors' success rate on the experimental cases (used as reference).	33% (8/24)	62% (15/24)	42% (10/24)	54% (13/24)	48% (46/96)
Full Performance Model	The success rate of the model that emulates the auditors' ascription errors is similar to the subjects' success rate.	42% (10/24)	71% (17/24)	54% (13/24)	67% (16/24)	58% (56/96)
Ablated Model (no bug B)	The success rate of the model that does not emulate the auditors' ascription errors (always ascribes) is higher.	75% (18/24)	79% (19/24)	92% (22/24)	87% (21/24)	83% (80/96)

The results in Table 3.6 are fairly uniform across cases, with the exception of Pharmaceuticals, which is characterized by a relatively small net effectiveness gain (from 71% to 79%). One reason for this outlier is a ceiling effect. The case elicited a larger percentage of successes than any other (about 20% more successes than the average of the other three). As a result, there was not much space for improvement when we ablated bug B. We also conducted an inspection of the protocols of the subjects who failed to detect the manipulations in the Pharmaceuticals case. The inspection revealed that these subjects frequently made hypothesis evaluation errors (error C) and evidence combination errors (error D), which we assume indicate the activity of bugs C and D. Ablating bug B in the presence of these other bugs is not sufficient to improve effectiveness.

Overall, the results of the comparison of the full and ablated performance models support our general argument: (1) ascription is a major determinant of success in a task that requires managing the risk posed by Others, and (2) ascription errors are a major component of the behavioral variance of subjects engaged in a task that requires the management of risk posed by Others.

The results of the auditing data also allow us to strengthen and qualify the conclusions reached from the analysis of the currency traders. Overall, the auditor data show that ignoring ascription behavior (errors B) increases the model's effectiveness and degrades its predictive power. The ablated model of second partner review is more effective because it does not make ascription errors. It can ascribe goals and actions to management and conclude that they have manipulated the financial information. At the same time, as the model gets more successful, its ability to approximate subjects' outcomes accurately worsens, which is consistent with the results obtained in the experiments with the traders' models.

Consistent with the previous analysis, we found that the auditors often failed to (abstain from) ascribe beliefs to the Other. Every subject in our sample made at least one ascription error, and some made as many as eight out of eight opportunities to err. Ascription errors were the most frequently observed type of error in the Johnson et al. (2001) study.

The last conclusion from the analysis of the traders was that ascription is not always an optimal behavior. Although the nature of the second partner review task requires the ascription of the intent to deceive, and therefore a failure to ascribe is tantamount to an error in this domain, we have observed a small number of individuals (3 out of 24) who succeeded at the task despite the fact that their behavior did not show evidence of ascription to management of the intent to deceive. These individuals felt that the cases contained (unintentional) mistakes of such magnitude that the case as a whole was misleading.

Conclusion

In spite of profound differences, spot currency trading and second partner review share a defining characteristic: the management of risk posed by others. Both traders and auditors operate in settings where the actions of Others create the possibility for loss. Currency traders manage the risks of holding large sums of foreign currency. Auditors manage

the risks of providing assurance that published financial information is free from deceptive manipulations.

We have highlighted two common features of these settings: dynamic interaction with many agents and asymmetric distribution of information under conflict of interest. Under either condition, individuals managing social exchange risk need to understand and predict the actions of others so as to take action to prevent or minimize losses.

Spot currency traders interact in highly dynamic markets that respond to the vagaries of global socioeconomic conditions and political uncertainty. The markets are too large for any one individual to control, and prices can change more rapidly than the trader can respond. Volatility in spot prices creates the possibilities for loss (i.e., risk) or, conversely, opportunities for profit. The highly dynamic nature of the spot markets is the overriding constraint on spot traders' decision making. Understanding and predicting the direction of the market is critical for trading success.

Second partner reviewers do not suffer from time pressure, yet have an equally challenging task: to detect manipulations of financial information that have been designed with the explicit intent to escape detection. Failure to identify these manipulations exposes the auditor to lawsuits and loss of professional reputation. Because the auditors do not have the same access to information that the management of a client company has, understanding the reporting goals of management is critical for auditing success.

Both traders and auditors are engaged in the social exchange of information. However, whereas fraud detection is an adversarial interaction, trading is not. Traders interact repeatedly, often several times a week. The countless iteration of one-on-one exchanges precludes deception and supports the development of trust. A dishonest trader soon has no trading partners and is out of business. In contrast, the auditors have to detect forms of deception: The management of a company in economic distress may deliberately choose to misstate the financial position of the company to obtain more favorable treatment from shareholders, investors, lenders, and regulators.

We have proposed that ascription, the ability to assign goals and beliefs to the Others with whom we interact, plays a major role in decision making in social exchange. We have supported our argument by means of anecdotal evidence from our empirical work and by means of a pair of experiments with two information processing models of the decision making required to manage social exchange risk. Comparison of

the process traces of the subjects' and systems' decision making suggests that the systems capture the basis for expert performance. Comparison across subjects and systems demonstrates that identification and processing of attribution-relevant categories mediate competent performance.

We found that experienced professionals involved in managing social exchange risk ascribe goals and beliefs to Others and that ascription is an important determinant of success. However, we also found that there are cases in which ascription does not lead to the best course of action, and that sometimes experienced professionals either do not ascribe or override the implications of ascribing goals and beliefs to Others. Overall, the role of ascription that emerges from these data is more complex than we initially expected.

Reckers et al. (1992) have observed that professional guidelines do not contain operational guidance about how to ascribe goals and beliefs to Others that pose risks. Perhaps some decision makers abstain from ascribing because they believe that ascription is too subjective or unreliable. There is a difference between ascribing and ascribing correctly, and the decisions based on ascription are hard to justify.

March (1988) has proposed that there are higher orders of ascription, which might have decreasing usefulness. We not only ascribe "first-order" beliefs and goals to Others, we also ascribe beliefs and goals that result from them ascribing beliefs and goals to us (see also Dennett, 1987). Indeed, the relationship between accountants and managers has been labeled as "a race to outwit each other." There is little doubt that higher-order ascription quickly becomes very complex.

More broadly, Dennett has argued that there is an evolutionary advantage for Agents who possess cognitive machinery that implements ascription. Recent work in cognitive psychology, neurology, and cognitive ethology supports Dennett's arguments (Baron-Cohen, 1995). We have examined evidence for the operation of such mechanisms in professional settings, and have argued that competent management of risk in social interactions derives part of its power from the cognitive ability to ascribe goals and beliefs to others.

If ascription is a deep cognitive capacity that promotes successful decision making in social settings, yet is not always applied, then there is need for research to better understand its use as well as to determine means to facilitate its exercise. Work by Cosmides, Tooby, and others has already shown that it is possible to elicit extraordinarily high success rates in simple experimental tasks by means of treatments rooted in

theories of reasoning about social exchange. What remains to be done is to generalize results obtained in tightly controlled laboratory conditions to more realistic decision-making situations that require managing risks posed by others.

Appendix A: The Currency Trading Model

This appendix describes the implementation of SPOT, the rule-based system that we developed to test our model of spot currency trading. Inputs to SPOT are database files, one for each of the 24 sessions in the laboratory experiment described earlier. The files summarize the events in the trader's experimental session: the news in the overnight report, the headlines the trader read during the trading day, the hourly prices posted on the price screen, and the requests for quotes the trader received. SPOT completes a trading day by cycling through all the lines in a file. The output that SPOT generates includes text files that detail the sizes of the positions it takes and the profits (losses) it creates.

As discussed earlier, SPOT contains methods for both feedback and feedforward process control. The method for feedback control compares the current rate of change in prices and the profits (losses) in a position to thresholds of acceptable risk. The method for feedforward control implements the process of ascription to anticipate the direction in which the market will drive prices whenever it "reads" a headline that the trader read in the experimental session.

Ascribing Beliefs to the Market

SPOT's representation of the market and its environment consists of three classes of objects – *Headlines*, *News Story*, and *World Events* – that constitute attribution-relevant cues and categories. Instances of the class *World Events* specify the algorithm for comparing situational cues either to a market standard or to prior cues when they exist. The comparison yields a value for the object *Market's View* of the appropriate instance of *News Story*. *Market's View of News Story* represents the product of SPOT's ascription of a belief to the market.

SPOT anticipates the direction in which the market will drive prices by reasoning about the beliefs the market forms about headlines. SPOT's limited domain knowledge about headlines (situational cues) is contained in 146 instances of a class of objects called *Headlines*. *Headlines*

contain two objects, *topic* and *category*. The values for these objects were set by the developer for each of the 146 instances of the class. The *topic of Headlines* for each instance represents a general type of event (e.g., budget surplus) rather than any particular event (e.g., Swiss budget surplus for February). Each instance is applicable to a large number of potential headlines. The *category of Headlines* for each instance links the associated topic to one of the 10 categories of news listed in Table 3.1. For instance, the *topic* WPI is linked to the *category* inflation.

CLASS Headlines
WITH topic STRING
WITH category COMPOUND
Economic growth,
Political uncertainty,
Inflation,
Central bank gov't policy,
Unemployment,
Old news,
Forecast,
Immaterial,
Up Trend,
Down Trend

The 146 instances capture all the knowledge about headlines that SPOT brings to a session. They represent what is undoubtedly a small fraction of the professional trader's knowledge about global socioeconomic and political events. They are the minimum number required to process the headlines in Smith's experiments and suffice to test the sufficiency of SPOT's method for anticipating the market.

The structure of the class *Headlines* finesses Trope's process of identification. The linkage between topics and categories provided by the class *Headlines* removes the need for the system to decide to which category of news each headline belongs. This finesse makes SPOT's identification as "spontaneous" as the traders'.

Whenever SPOT "reads" a headline (an entry in the input database), either in the overnight report or during the trading day, it generates or updates an instance of the class of objects called *News Story*. Instances of the class *News Story* represent SPOT's memory of headlines it has read. For instance, when SPOT read the overnight report for the WPI example, it generated an instance of *News Story* with the objects *city* set to Washington, *topic* set to WPI, *relevant currency* set to the dollar,

and *expected value* set to +0.4%. The object *expected value of News Story* represents the information the market was said to expect when it next sampled a related headline.

```
CLASS News Story
    WITH city STRING
    WITH topic STRING
    WITH relevant currency STRING
    WITH expected value NUMERIC
    WITH index NUMERIC
    WITH Market's View COMPOUND
        bullish,
        bearish,
        of little concern
    WITH relative to market standard COMPOUND
        below,
        above
```

The knowledge captured by objects in each instance of *News Story* represents all of SPOT's knowledge of prior cues.

SPOT's knowledge of attribute-relevant categories is contained in 10 instances of the class *World Events*. There is one instance for each category of news shown in Table 3.1. The knowledge captured by these 10 instances represents knowledge the trader brings to the trading desk. The values of the two objects, *comparison algorithm* and *market standard*, were assigned by the developer for each instance. The assignments follow the specification given in Table 3.1. For example, the *comparison algorithm* for the *category* inflation is *down is good* (i.e., a decrease in inflation is viewed as beneficial to the relevant currency) and the *market standard* is 2%. The object *market standard* was left blank in the seven instances where it does not apply.

```
CLASS World Events
    WITH category STRING
    WITH comparison algorithm COMPOUND
        Up is good,
        Down is good,
        Bullish,
        Bearish,
        Ignore
    WITH market standard NUMERIC
```

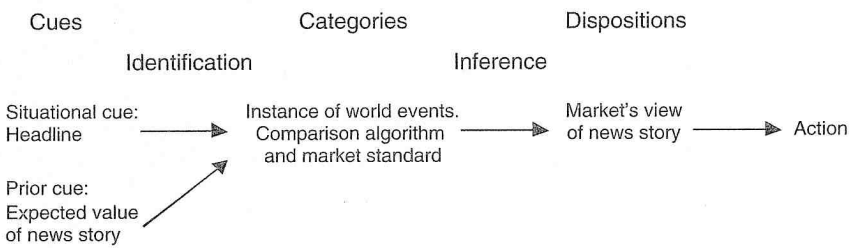


Figure 3.A1 SPOT's implementation of ascription when it finds a prior cue.

SPOT implements Trope's process of dispositional inference whenever it reads a headline either in the overnight report or during the trading day. It first seeks to match the *category* of the active instance of the class *Headlines* to the *category* of all existing instances of the class *News Story*. If a match is found (e.g., if SPOT had read the overnight report to indicate what the market was said to expect), it updates the *index of News Story* to the information contained in the headline. As shown in Figure 3.A1, SPOT then applies the *comparison algorithm* of the appropriate instance of *World Events* to compare the (new) *index of News Story* to the preexisting *expected value of News Story*. The comparison allows SPOT to attribute a belief to the market based on both the information in the headline and the market's expectation for the headline.

In the example, SPOT found an instance of *News Story* with *city* Washington and *topic* WPI. It updated *index of News Story* to 0.6%, found an *expected value* of 0.4%, and applied the *comparison algorithm of World Events* – down is good – to set the *Market's View of News Story* to *bearish*. The inference of a bearish market view captures SPOT's ascription to the market of the belief that increases in inflation are disadvantageous to the *relevant currency*, in this case, to the dollar.

The search for a matching instance is necessarily unsuccessful the first time SPOT encounters a particular headline during a session (e.g., when it samples the overnight report). Figure 3.A2 shows how SPOT attributes a belief in the absence of a prior cue. When the search is unsuccessful, the system creates a new instance of the class *News Story*. It then matches the *category* of the active instance of the class *Headlines* to the *category* of the 10 instances of the class *World Events* and applies the associated comparison algorithm to compare the (new) *index of News Story* to the preexisting *market standard of World Events*. The

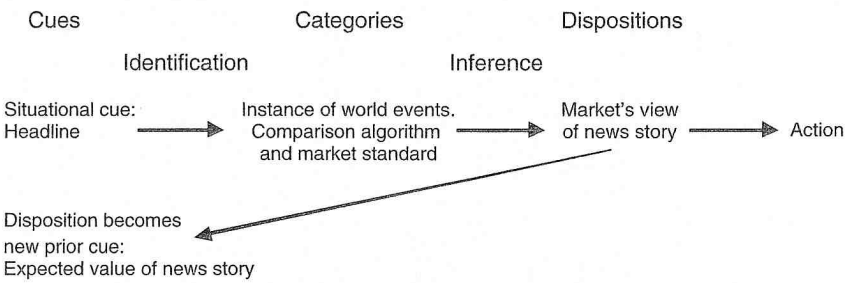


Figure 3.A2 SPOT's implementation of ascription when it does not find a prior cue.

comparison allows SPOT to attribute a belief to the market based on the information in the headline and the default standard for that class of headline.

The result of the comparison is flagged by the object *Market's View of News Story*. The comparison may be either bullish, bearish, or of little concern for the relevant currency. The comparison algorithm for headlines that fall in the categories of news *immaterial, old news*, and *forecasts* specifies that the *Market's View of News Story* is of little concern.

The beliefs ascribed to the market are contextually dependent on the presence or absence of a prior cue. For instance, in the absence of a prior cue, a headline that suggests that growth will remain moderate triggers a standard for comparison that supports the inference that the market's response will be tepid. The same headline, in the context of a prior cue that the market expects news of robust growth, triggers a comparison that supports the inference that the market's response will be draconian and swift. Distinguishing between situational and prior cues enables SPOT to capture the contextual sensitivity to expectations that can be critical to success in anticipating the market.

Appendix B: The Model for Detecting Financial Manipulations

To evaluate our model for detecting financial manipulation, we implemented it as a computer program. This appendix provides details on the algorithms contained in the model and on its implementation as a rule-based system. The inputs to the model are the financial statements of the companies included in the experimental materials. The model's output is a diagnostic opinion qualifying the extent to which the

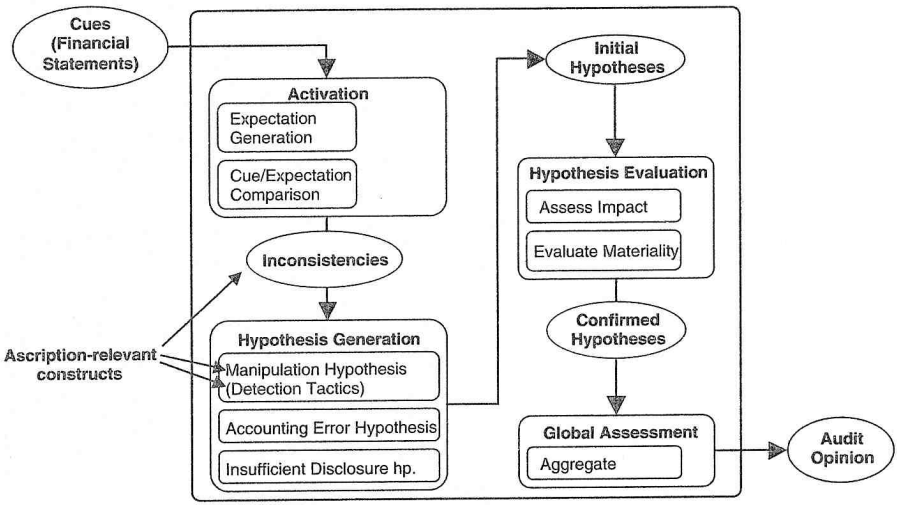


Figure 3.B1 Process model of the detection of financial deception. (Adapted from Johnson et al., 1998.)

financial statements are a fair representation of the company's economic condition (e.g., "the statements are misleading" or "the statements are fair"). The model is based on the assumption that the detection of financial manipulation can be decomposed into four distinct processes: activation, hypothesis generation, hypothesis evaluation, and global assessment.

Activation

The activation process generates expectations for cues in the financial statements, compares each cue with its expectation, and, based on the comparison, flags them as inconsistent (see Figure 3.B1). Expectations are produced by simple linear functions of past values of the target cue, other cues, and industry-characteristic constants (Johnson et al., 1992). As an example, the expected value of inventory is computed by multiplying the past value of inventory by the rate of growth in sales. In the previous protocol excerpt, S17 used a similar technique: He computed the rate of growth in inventory and compared it to a standard (the rate of growth in sales).

Whenever the difference between a cue and its expectation is larger than a specified criterion value (e.g., 2.5% of operating income), the cue is labeled as inconsistent. Inconsistencies (e.g., "the inventory level is too

high") are symptoms of possible anomalies in the accounting process that produced the financial statements. Inconsistencies include what Trope calls *ascription-relevant categories*.

Hypothesis Generation and Ascription

Hypothesis generation is the process that implements ascription in the model. The process generates initial hypotheses that explain the inconsistencies detected in the financial statements. Three classes of hypotheses can be generated: The anomaly results (1) from intentional manipulations of the accounting process, (2) from unintentional mistakes in accounting, or (3) from insufficient disclosure of information.

The generation process is activated by the detection of an inconsistency (see Figure 3.B1). First, it attempts to apply the detection tactics to it. If a tactic applies, an hypothesis of intentional manipulation is generated for the inconsistent cue. If no tactic applies, then either an insufficient disclosure hypothesis or an accounting mistake hypothesis is generated. An insufficient disclosure hypothesis is generated when the inconsistent cue is related to a note to the financial statements (e.g., a change in accounting policies). Otherwise, an accounting mistake hypothesis is generated for the cue.

The tactics are fired when three ascription-relevant categories are present: (1) a cue is flagged as inconsistent with its expectation, (2) the inconsistent cue is functional to a possible goal of management, and (3) the inconsistency is feasible, that is, it might have been caused by management. The first two conditions are common to all tactics, and the third distinguishes among them. To establish functionality, the model provisionally ascribes to management the general goal of increasing earnings (similarly to what S17 did) and then examines the cue flagged as inconsistent. If the inconsistent cue is an asset (a liability) and is larger (smaller) than expected, then the model concludes that the inconsistency is functional to the ascribed goal.

The second condition – feasibility – is established when a manipulation of the accounting process that explains the inconsistency is available. The model contains about a dozen manipulations derived from published lists (e.g., Schilit, 1993). Manipulations are associated with specific cues based on each cue's position in the financial statements (e.g., as an asset, an expense, and the like). For instance, capitalized tooling costs are associated with repackaging. The output of the hypothesis generation process is an initial (i.e., subject to evaluation) set of

hypotheses. These hypotheses are the input to the next subprocess that evaluates them.

Hypothesis Evaluation

The initial hypotheses generated by the model need to be confirmed. A process of hypothesis evaluation produces confirmed hypotheses on the basis of the relevance (*materiality* in accounting terminology) of the inconsistency associated with each hypothesis.

The type of hypothesis proposed for an inconsistency suggests the nature of the test for materiality of the inconsistency. For instance, insufficient disclosure hypotheses are evaluated on the basis of the magnitude (a dollar value) of the difference between the cue value and its expectation. The repackaging hypothesis prescribes recategorizing the entity that may have been miscategorized under a worst-case scenario assumption. After these dollar values are assigned, the initial hypotheses are individually evaluated against an assumed criterion for materiality (10% of the operating income; Johnson et al., 1992).

Global Assessment

The global assessment process aggregates and evaluates the confirmed hypotheses to produce a final outcome for the second partner review task. The model selects one of three possible outcomes (Grazioli, 1997; Johnson et al., 2001): "Misleading", "Unqualified with additional paragraph," and "Unqualified."

Misleading expresses the conclusion that the reviewed financial statements are an unfair representation of the company's underlying economic conditions. Misleading is the correct solution for all the tasks presented here. Unqualified with additional paragraph means that the model identified some audit issues (e.g., lack of consistency, going concern) in a case, yet these issues are not sufficient for concluding that the financial statements are misleading. An opinion of Unqualified is an indication that the model did not identify any issue or manipulation in a case. This is the outcome desired by the company's management.

Outcome selection is based on either one hypothesis or two or more combined hypotheses. If an individual manipulation hypothesis is material, then the global assessment process selects a Misleading outcome. If a mistake in accounting or an insufficient disclosure hypothesis is

material, then the process selects an Unqualified with additional paragraph outcome. Otherwise, an Unqualified outcome is selected. Aggregate evaluation is performed by first cumulating the dollar values associated with each type of hypothesis, and then testing these values against a criterion for aggregate materiality (40% of the operating income). If any aggregate hypothesis is material, the global assessment process selects a Misleading outcome.

References

Albrecht, W. S., Wernz, G. W., & Williams, T. L. (1995). *Fraud: Bringing light to the dark side of business*. New York: Irwin.

Bainbridge, L. (1974). Analysis of verbal protocols from a process control task. In E. Edwards & F. P. Lees (Eds.), *The human operator in process control* (pp. 146–158). London: Taylor & Francis.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and the theory of mind*. Cambridge, MA: MIT Press.

Bowyer, J. B. (1982). *Cheating*. New York: St. Martin's Press.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*, 211–241.

Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology, 29*, 181–201.

Ceci, S. J., Leichtman, M. D., & Putnick, M. E. (1992). *Cognitive and social factors in early deception*. Hillsdale, NJ: Erlbaum.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391–416.

Clark, A. (1987). From folk psychology to naïve psychology. *Cognitive Science, 11*, 139–154.

Cosmides, L. (1985). *Deduction or Darwinian algorithms? An explanation of the "elusive" content effect on the Wason selection task*. Ph.D. dissertation, Harvard University.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptation for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

Cosmides, L., & Tooby, J. (1995). Foreword. In S. Baron-Cohen, *Mindblindness: An essay on autism and the theory of mind* (pp. xi–xvii). Cambridge, MA: MIT Press.

Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Edwards, E., & Lees, F. (1974). *The human operator in process control*. London: Taylor & Francis.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99*, 3–19.

Ekman, P. (1991). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: W. W. Norton.

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise*. New York: Cambridge University Press.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.

Grazioli, S. (1997). *Detecting manipulations of financial information*. Ph.D. dissertation, University of Minnesota.

Grazioli, S. (1998). Errors in detecting financial deception: A cognitive modeling approach. *Proceedings of the Fourth Conference of the Association for Information Systems*, 152–154.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Holland, J. H. (1975/1992). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

Johnson, P. E., Grazioli, S., & Jamal, K. (1993). Fraud detection: Deception and intentionality in cognition. *Accounting, Organization and Society, 18*(5), 467–488.

Johnson, P. E., Grazioli, S., Jamal, K., & Berryman, G. R. (2001). Detecting deception: Adversarial problem solving in a low base rate world. *Cognitive Science, 25*(3), 355–392.

Johnson, P. E., Grazioli, S., Jamal, K., & Zualkernan, I. A. (1992). Success and failure in expert reasoning. *Journal of Organizational Behavior and Human Decision Processes, 53*(2), 173–203.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*, 1–24.

Kelley, H. H. (1971). *Attribution in social interaction*. Morristown, NJ: General Learning Press.

Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan.

Kruglanski, A. W. (1980). Lay epistemological processes and contents: Another look at attribution theory. *Psychological Review, 87*, 70–87.

Kurtzman, J. (1993). *The death of money*. Boston: Little, Brown.

March, J. G. (1998). *Decisions and organizations*. Oxford: Basil Blackwell.

March, J. G., & Shapira, Z. (1987). Managerial perspectives on risk and risk taking. *Management Science, 33*(11), 1404–1418.

Mitchell, W., & Thompson, N. S. (1986). *Deception: Perspectives on human and non-human deceit*. Albany: SUNY Press.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, Chap. 40, pp. 1–51). New York: Wiley.

Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18*, 87–127.

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*, (Vol. 2, Chap. 43, pp. 1–39). New York: Wiley.

Peters, J. M. (1990). A cognitive computational model of risk hypothesis generation. *The Journal of Accounting Research, 28*, 83–109.

Reckers, P. M. J., Wong-On-Wing, B., & Krull, G. W. (1992). Auditors' assessment of management's disposition: An attributional analysis. *Decision Science, 23*, 957–972.

Schilit, H. M. (1993). *Financial shenanigans: How to detect accounting gimmicks and fraud in financial reports*. New York: McGraw-Hill.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*(2), 252–266.

Shapira, Z. (1995). *Risk taking: A managerial perspective.* New York: Russell Sage Foundation.

Shoemaker, P. J. H. (1993). Determinants of risk-taking: Behavioral and economic views. *Journal of Risk and Uncertainty, 6,* 49–73.

Smith, K. C. S. (1996). *Decision making in dynamic environments: Trading in the spot currency markets.* Ph.D. dissertation, University of Minnesota.

Thagard, P. (1992). Adversarial problem solving: Modeling an opponent using explanatory coherence. *Cognitive Science, 16,* 123–149.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93*(3), 239–257.

Trope, Y., Cohen, O., & Maoz, Y. (1988). The perceptual and inferential effects of situational inducements on dispositional attribution. *Journal of Personality and Social Psychology, 55*(2), 165–177.

Van Lehn, K. (1990). *Mind bugs: The origins of procedural misconceptions.* Cambridge, MA: MIT Press.

von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Wiener, E. L. (1984). Vigilance and inspection. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 207–246). New York: Wiley.

Yates, J. F., & Stone, E. R. (1992). The risk construct. In J. F. Yates (Ed.), *Risk-taking behavior* (pp. 1–25). New York: Wiley.

Running head: SIMULATED TOUCH STIMULATES TRUST

Simulated Touch:

A Method for Stimulating Trust in Remote Command and Control

Kip Smith

Linköping Institute of Technology, Sweden

Martin Liberg and Jens Hilmersson

University of Skövde, Sweden

Rebecca Sitzes

Kansas State University

Abstract

This paper discusses an experiment that tested whether augmenting verbal orders with remote-controlled tactile stimulation would increase the effectiveness of remote command and control. Remote command and control entails the physical separation of a leader and those who are led and reliance on electronically mediated communication to give and receive orders. We have previously shown that verbal orders are obeyed more slowly when given by a remote leader than by a leader who is physically present, and that the delay is due to the lack of leader presence and not to reliance upon mediated communication. Here we document that remote-controlled tactile stimulation by a remote leader compensated for the previously observed delay in response times to remotely-issued verbal orders.

Simulated Touch:

A Method for Stimulating Trust in Remote Command and Control

Establishing trust among members of distributed teams is a pressing issue for industry (Meyerson, Weick, & Kramer, 1966; Olson & Olson, 2000) and the military (Adelson, 1961; Kaempf, Klein, Throdsen, & Wolf, 1996; Johnson & Levis, 1988). Managerial research on distributed work has tended to focus on the impact and acceptance of the technology that enables it rather than the psychological processes and states that sustain it (but see DeSanctis & Monge, 1999). In contrast, military research has tended to focus on the effectiveness of leadership from a distance and of the organizational structure known as *remote command and control* (Smith & Pangburn, in review).

We begin this paper with discussions of remote command and control, the construct of trust, and the negative implications of remote command and control on a team's trust in its leader. We then discuss remote-controlled tactile stimulation as a leadership tool for creating a bond of trust between individuals who have no opportunity for direct, physical contact and who must communicate electronically. We then present an experiment that tests whether remote-controlled, simulated touch facilitates remote command and control. The context of the experiment is military. The technology that allows us to conduct an ethical experiment with human participants while maintaining sufficient ecological validity to generalize to a military setting is called paintball. We use a paintball assault lane to test the utility of remote-controlled tactile stimulation as a method for enhancing soldier performance under fire. We argue that trust between members of a team and its leader is the mediating factor that links a remote leader's simulated touch and team performance.

Remote command and control

Remote command and control is an organizational structure with four defining features. First, there is a leader or group of leaders in one location and one or more work teams in geographically remote locations. Second, due to their physical separation, all communication between the leader and the team is mediated by telecommunications and other forms of technology. Third, the environment in which the team works is typically highly dynamic and rarely predictable (Brehmer, 1992). Indeed, environmental complexity and uncertainty are the major desiderata for adopting remote command and control (Adelson, 1961; Brehmer & Sundin, 2000). Finally, the organization is patently hierarchic rather than laterally connected. Leaders have a level of authority that the work teams recognize and cannot or do not challenge. For example, in the military, the remote leader is an officer who, among other tasks, issues orders. The team consists of recruits or reservists whose job it is to obey those orders. In industry, the formal power is rarely as absolute as it is in the military but the hierarchic organization is often strong.

Remote command and control has several advantages over structures that collocate the leader and the team. In industry and the classroom, it enables employees and students to work and take classes where they live. It makes it easier to change leaders (teachers) as the work evolves. In the military and some industries, remote command and control insulates the (expensive, knowledgeable, veteran) leaders from hazardous, uncertain work environments. In addition, a good command and control center is designed to provide the leader(s) access to multiple sources of information that are not routinely available at the work site or on the battlefield (e.g., Friman, 2000). The centralization of decision making and of the flow of

information facilitates the selective deployment of limited resources and speeds the cycle of decision making.

In previous work we have examined the effect of remote command and control and electronically mediated communication on the efficiency with which a remote leader's verbal orders are followed (Pangburn, Freund, Pangburn, & Smith, 2003; Smith, in press). Experimental participants needed more time to respond to a verbal order issued by a remote leader than by a collocated leader. The delay is due to the lack of leader presence and not to reliance upon electronically-mediated communication (Smith & Pangburn, in review). In the work discussed here, we examine an effort to mitigate this unfortunate consequence of remote command and control.

Research on remote command and control has a colorful tradition. In classic experiments on the 'obedience problem,' Stanley Milgram found that a leader's authority over a subordinate was a function of the leader's proximity (Milgram, 1974). The participants in Milgram's research were three times more likely to administer electric shocks to another person when instructed by a leader to do so when the leader was present than when the leader gave the order from a separate room via telephone. Milgram's finding that collocation is instrumental to obedience is highly relevant to the work discussed here since both are intended to generalize to the bond between a officer and a soldier in battlefield situations.

Trust

Following Rousseau, Sitkin, Burt, and Camerer (1998), we define trust as the willingness of one party to be vulnerable to the actions of another based upon positive expectations of the other's behavior. The critical elements in this definition are risk and interdependence. Risk is a

psychological state. It is awareness of an opportunity for meaningful loss (MacCrimmon & Wehrung, 1985; Yates & Stone, 1992). Interdependence is an environmental state. It is the state of affairs in which two or more individuals need and seek to rely on one another to achieve or maintain a common goal. Trust emerges if and only if individuals acknowledge their interdependence with respect to a desired risky outcome. Trust makes individuals willing to form relationships that take risks and to take risks in a relationship (Mayer, Davis, & Schoorman, 1995).

Trust is not a behavior, rather it is a psychological state that influences the interaction between two or more individuals or groups (Berg, Dickhaut, & McCabe, 1995; Hoffman, McCabe, & Smith, 1998). However, behavior and trust are closely related. Trust is partially based upon an individual's observations of another's actions (Whitener, Brodt, Korsgaard, & Werner, 1995). For example, behaving in a trustworthy manner towards another person may mark the beginning of a trusting relationship (Meyerson, Weick, & Kramer, 1996).

Trust has been closely linked to higher levels of performance (Costas, Roe, & Taillieu, 2001) and to cooperative behaviors in organizations (Mayer, Davis, & Schoorman, 1995). Workers who feel that they can trust their organization and their leaders are more likely to cooperate with others within their organization, as well as to engage in extra-role behaviors (Lewicki, & Bunker, 1996).

McAllister (1995) posits that trust has both emotional and cognitive foundations. Cognition-based trust is thought to be rooted in professionalism, integrity, trustworthiness, and reliability. Initial levels of cognition-based trust between individuals who never have met must be built from declarative knowledge about each other's professionalism and previous

achievements. Meyerson, Weick, & Kramer (1966) propose that members of a virtual team can rely on each other's reputations for professionalism to develop cognition-based trust quickly. Whether based on reputation or experience, cognition-based trust can be summarily torn down by, for example, a single ill-advised decision or bungled mission (Turner, Hogg, Turner, & Smith, 1984).

In contrast, affect-based trust takes time to develop (McAllister, 1995; Kanawattanachai & Yoo, 2002). Affect-based trust is rooted in visceral connection, consideration, and empathy and is informed by social skills and affective traits such as care and thoughtfulness. It can be built upon previously-formed cognition-based trust (McAllister, 1995), mutual experience (Jones & George, 1998), or informal conversation (Orbell, Dawes, & van de Kragt, 1988), but may be unlikely to form over the telephone or two-way radio (Rocco, 1998). In the absence of physical proximity and familiarity, levels of affect-based trust are likely to be low.

While both affect-based and cognitive-based trust contribute to the construct of trust, they have differential impact across situations. Affect-based trust is more prominent in social or family life. Cognition-based trust is more salient in the workplace. Research on distributed work in virtual environments has shown the value of cognition-based trust (Handy, 1995; Jarvenpaa & Leidner, 1999). Collaboration that establishes cognition-based trust is more likely to be successful even when conducted over electronically-mediated communication networks (Kanawattanachai & Yoo, 2002).

The consistency of an individual's behavior is an important component in the maintenance of both affect- and cognition-based trust. People who can be relied upon to behave in a trustworthy way will likely be trusted (Whitener et al., 1995). Within a work group, the

mutual need for trust is such that a high level of cognition-based trust can lead to the growth of affect-based trust (Jones & George, 1998; McAlister, 1995).

These considerations make it reasonable to believe that a work team's levels of affect-based and cognition-based trust in their leader is a function of their experience working together (e.g., Dirks, 2000; Wageman, 1995). The tradition of having soldiers and their officers develop camaraderie and fellowship by training together and working side-by-side on the battlefield appears ideally suited to the development of cognition-based and, in time, affect-based trust.

Network-based Defense

The domain of application of our study is a proposed redesign of the organizational structure of military units of four to eight foot soldiers known as *fire teams*. A fire team is the basic fighting unit in the dismounted infantry. The U. S. Army and the Swedish Ministry of Defense have started a process that will change how fire teams are formed and how they are led (Brehmer & Sundin, 2000; Friman, 2000; Hébert, 2003). The proposed redesign is called the *objective force warrior* in the USA and *network-based defense* in Sweden. We summarize how fire teams have traditionally been formed and led before we discuss the redesign.

Traditionally, a fire team and its officer develop bonds of mutual trust by sharing an intense and arduous regimen of training that leads them to care for each others' welfare. On the battlefield, the fire-team leader has traditionally received mission directives and been charged with assessing the evolving situation and directing the team to accomplish the mission. Because the fire team and its leader train and fight side-by-side, orders and other communication have been mediated by speech, gesture, and physical contact.

This time-honored approach to the development of trust and camaraderie in fire teams may become part of military history (Eden, 1986; Sundin & Friman, 2000). The objective force warrior will receive specialized training in the conduct of a specific battlefield task (Throne, 2003). Consequently, fire teams may be assembled by selecting individual soldiers according to their specialties and depending on the mission. The resulting fire team is likely to have no or relatively little common experience and training.

A second part of the redesign is the adoption of remote command and control of fire teams. Soldiers on the battlefield will no longer take their orders from an officer whom they can see. Instead, orders will be given by a leader located in a safe position away from the battlefield. For example, an American fire team on a mission outside American territory may be led by an officer sitting in The Pentagon.

One consequence of this fundamental change in team dynamics is that the members of a fire team may know their remote leader or they may not. It is reasonable to believe that the formation of trust will be difficult in an organizational setting like remote command and control that precludes familiarity and proscribes physical contact. A second consequence is that all communication between soldiers and officers must be mediated by radios or other portable communications equipment. Accordingly, part of the vision for the objective force warrior is to require all soldiers to carry various types of sensors and telemetry equipment so that the remote leader can see what they see (Hébert, 2003).

A third and potentially devastating consequence of remote command and control is loss of physical proximity between a fire team and its officer on the battlefield (Smith, in press). The soldiers will be exposed to live fire while the officer remains out of harms' way. Removing the

fire-team leader from the field of battle eliminates the mutual vulnerability that is a necessary precondition for the formation of trust (e.g., Rousseau et al., 1998). This makes it highly unlikely that a remotely-led fire team will be able to develop any form of affect-based trust (McAllister, 1995). Initially, camaraderie within the fire team and their trust in their remote leader will have to rely on some form of cognition-based trust.

In sum, the vision for the reorganization of the fire team is likely to reduce the fire team's trust in its leader in two ways. First, by assembling teams composed of soldiers who are likely to be strangers, the vision undermines camaraderie and affect-based trust. Second, the reliance on remote command and control eliminates all but the leader's reputation or rank as the basis for developing cognition-based trust. This analysis suggests that the proposed organizational change is likely to undermine the trust between a fire team and its officer that for millennia has been the foundation of military effectiveness.

Given the U. S. and Swedish militaries' commitment to remote command and control and its growth in both industry and pedagogy (distance education), there is a need to develop and test methods for the in-situ formation of trust between unfamiliar individuals.

Touch

Physical contact has been seen to create a basis for increased compliance to a request (Guéguen, 2004). Working in a classroom setting, Guéguen contrasted two conditions. In the control condition, the teacher gave positive reinforcement to students individually. In the touch condition, the teacher lightly touched the students on the upper arm for 1 second while giving positive reinforcement individually. At the end of the class, a higher percentage of students in the touch condition volunteered to go to the blackboard and show their work.

Guéguen proposes complementary explanations for the effect of touching and being touched on behavior. The first involves social status. A touch from a person with relatively high social status (e.g., a teacher or an officer) induces confidence and a sense of value in those with relatively low social status. Guéguen's second explanation deals with the activation of moods. When socially and culturally appropriate, being touched by another person has been seen to induce a positive mood in those who are touched (Henley, 1973). In Guéguen's classrooms, feelings of confidence, validation, or an associated positive mood may have made the students more willing to demonstrate their work in front of the rest of the class. Hall (1996) has shown that (in western cultures) supportive touch has its greatest impact on performance when applied to that part of the back or to the upper arm.

We suggest that a culturally-appropriate touch from a person of higher status communicates a willingness to establish interdependence and to provide support. Thus, touch not only leads to the affective benefits posited by Guéguen but also is the impetus for the dawning of cognition-based trust. Touch may provide a mechanism for quickly establishing trust between a team leader and members of a team.

Issues and Hypothesis

In previous work we found that experimental participants needed more time to respond to a verbal order issued by a remote leader than by a collocated leader (Pangburn, et al., 2003; Smith, in press). The delay is due to the lack of leader presence and not to reliance upon electronically-mediated communication (Smith & Pangburn, in review). We interpret these results to suggest that a lack of trust in the leader is the intervening factor that induces the delay in the execution of orders under remote and command and control. Here we report a test of

whether remote-controlled, simulated touch may compensate in part for the delay and, by inference, facilitate the formation of a bond of trust.

The context of the investigation is a fire-team and its remote leader. Participants played the roles of soldiers and snipers in a paintball assault lane. An experimenter played the role of the remote leader. We expect participants in the role of soldier to respond more quickly to verbal orders issued over a radio by the remote leader when the orders are accompanied by a tactile stimulus than when they are not accompanied by a tactile stimulus.

Method

The method used in this experiment is largely based on Smith's (in press) experiments that investigated the influence of leader presence in stressful conditions. The experiment was conducted in Swedish by two of the junior authors.

Participants

Some of the experimental participants were secured through personal contact by the experimenters, some through notices posted on a bulletin board at the University of Skövde, and some by email broadcast. A total of 26 people took part in the experiment, 24 men and 2 women. Ages ranged from 19 to 27 years with a median of 22. Eight of the men and one of the women had previous military experience lasting between 3 and 12 months. Twenty-three were Swedish, two were Icelandic, and one was from the Indian subcontinent. Twenty-five spoke fluent Swedish.

All volunteers were warned, both verbally and in writing, that they would be participating in an exercise that involved the fire of live, non-lethal ammunition (paintballs). They were

further told, both verbally and in writing, that they were likely to experience a moderate amount of pain in the course of the experiment but no more pain than they would usually experience playing recreational paintball. They were also told, both verbally and in writing, that they would be called upon to shoot at another person. All agreed to these conditions. These efforts allow us to assert that we treated the participants ethically.

The results from two of the participants were not included. The response-time data from one of the participants were judged to be extreme (more than three standard deviations longer than those of the other participants). The data from another participant was highly erratic. This behavior may have been induced by confusion associated with lack of fluency in Swedish. Of the remaining 24 participants, 12 played the role of soldier in the control condition first and 12 played the role of soldier in the touch condition first in a repeated-measures design.

Experimental setting and apparatus

Layout. The experiment was conducted inside a 40 x 16 m paintball arena in Tidan, 15 km north of Skövde Sweden. As shown in Figure 1, the assault lane was delineated by eight large (2.5 m long by 1.5 m high) inflatable protective barriers, four on each side, 8 m apart. A small cup containing five paintballs was placed behind each barrier. At the far end of the lane stood a protective wall containing two horizontal slots (0.5 m long by 0.2 m high) at shoulder height. The wall formed the sniper's hardened position. The slots allowed the sniper to shoot at (and be shot by) the soldier. A table made of shipping pallets was placed directly in front of the slots in the sniper's bunker. Five targets, 1.5 liter plastic bottles filled with colored water, four red and one blue, were placed in random order in a row on the table.

An observation room above and behind the assault lane contained two screened windows which provided a commanding view of the lane, the soldier, and the sniper below. The remote leader and the observer stood in this room away from the windows to avoid being inadvertently seen by the participant playing the role of soldier.

Munitions and communications. The paintball rifle was the armament used in the experiment. Ammunition consisted of either green or yellow paintballs. Participants, both the soldier and the sniper, were issued protective clothing consisting of a face shield and coveralls. The communication equipment used to issue verbal orders were Danita SP3380 walkie-talkies equipped with and voice-activated by headsets.

Tactile stimuli. The remote-controlled vibrator package, shown in Figure 2, was assembled by an undergraduate research assistant at the University of Skövde. The conceptually elegant equipment was cheap, low-tech, and error-free. It had two parts, a tactile-stimulus unit and a communications unit. The participant playing the role of soldier wore the package in both conditions of the repeated-measures design.

The communication unit consisted of two walkie-talkies, tuned to the same frequency, and a sound-activated relay. One of the walkie-talkies was the remote leader's sending unit and the other the soldier's receiving unit. The leader initiated a tactile stimulus by pushing the 'call' button on the sending unit. This signal was received by the soldier's walkie-talkie and sent electronically to the sound-activated relay. The receiving unit and the relay were worn on a waist belt by the soldier. The relay was, in turn, connected by wire to the tactile-stimulus unit.

The tactile stimulus was a mild buzzing rumble that lasted for 2 seconds. The stimuli were delivered by a small metal box containing an electric motor driving an eccentric cam. The

box was fastened by a belt that wrapped around the shoulders and ribcage of the participant playing the role of soldier so that the box sat between the shoulder blades. The brief buzzing on the back was intended to simulate tactile communication by a collocated leader and to infuse the trust that we propose touch can convey.

The off-the-shelf hardware of the remote-controlled vibrator package made it necessary for the leader to activate two walkie-talkies simultaneously in the touch condition. (The two sets of walkie-talkies, one for issuing verbal commands and the other for activating the tactile stimulus unit, used different frequencies.) This made it difficult for the leader to issue verbal orders and administer tactile stimuli in one stoke. Accordingly, it is unlikely that all tactile stimuli in the touch condition were received at exactly the same time as the corresponding verbal orders. However, any lack of synchrony between the two types of signal was likely to be randomly distributed within and across participants. We are working on ways to improve the design of our low-tech solution to this high-tech problem.

Design

To test whether remote-controlled touch can simulate leader presence and, as a consequence, can compensate in part for the significant time delay found by Smith (in press), this experiment used two conditions of leader presence in a repeated-measures design. In the control condition, a remote leader issued orders using a radio link; in the touch condition, the leader activated the tactile-stimulus unit at the same time as he issued an order. There were two types of orders, to move (advance to the next barrier in the lane) and to shoot (at the targets in front of the sniper's bunker).

The sequence of the two conditions, with and without remotely controlled tactile stimulation, was randomly assigned across participants. Since we have already documented a contrast in response times to orders given over a radio by a collocated and by a remote leader, we did not include a condition with a collocated leader.

Tasks

Soldier. Every participant played two roles, soldier and sniper, in random order. The task when playing the role of soldier was to move forward from barrier to barrier and to fire at the targets upon receiving orders to do so. Each participant completed the mission two times in the role of soldier. As far as the soldier's mission was concerned, the touch and control trials were identical with the exception of the presence of tactile stimuli in the touch condition. In both conditions, the soldier used a walkie-talkie equipped with a headset to communicate verbally with the remote leader.

Sniper. A second participant acted as the sniper. The sniper's task was to shoot the soldier, that is, to instill anxiety in the soldier and to slow his or her advance up the assault lane. The opportunity to play the role of sniper was a major inducement to participation.

The sniper's bunker consisted of a protective wall at the end of the assault lane. The wall contained two slots that provided a clear view of the lane and insured that the solder would always be in the sniper's field of fire. The sniper was directed, by the leader in the presence of the soldier, to pause briefly when moving from one of the slots in the wall to the other. The reason for directing the sniper to move slowly between the slots was to give the leader grounds for judging when it would relatively safe for the soldier to move and to shoot. Both the soldier and the sniper were asked to explain the implications of these instructions to the leader before the

war game started. The sniper was then issued 40 paintballs and asked to man the position behind the protective wall.

Leader. One of the experimenters, who had military experience, played the role of leader in every session and condition throughout the experiment. The leader stood in the observation room above and behind the assault lane out of the soldier's view, Figure 1. From there the leader issued verbal orders to the participant playing the role of soldier to move and to shoot. The timing of these orders was based upon the position of the sniper and the leader's judgment of when it would be relatively safe to move or to open fire. The orders that were given in both conditions were "fire" ("eld") and "advance" ("framåt"). In the touch condition, every order was accompanied a remote-controlled tactile stimulus. The leader pressed the call button on the walkie-talkie that activated the vibrator motor at the same time as he pressed the call button on the walkie-talkie used to issue the order.

Procedure

Participants were picked up in groups of three at the University of Skövde and taken by car to the indoor paintball arena in Tidan. During the twenty-minute ride, they were given and asked to read a paper copy of a mission brief that described a fictitious scenario which provided the background for the task they would be asked to perform. The English translation of the mission brief is included in the appendix. They were also given an informed consent form which they were asked to read and sign and were encouraged to ask questions. The consent form advised them about the possibility of incurring some pain, the need to shoot paintballs at another person, and the fact that the research is supported by the Army Research Office of the United States of America.

On arrival in the paintball arena, participants were issued protective coveralls and a facemask. They were outfitted with the remote-controlled vibrator equipment which was worn under the coveralls and were introduced to the sensation of the tactile stimulus. A member of the club that runs the arena instructed them on the use and safety features of paintball rifles.

Participants then met their (male) leader for the first time. The leader was dressed in military fatigues and had a military bearing. The leader told them he would be giving them two types of orders, (1) to advance up the assault lane from one barrier to the next and (2) to shoot at the targets in front of the sniper's bunker. He instructed them to obey all orders as quickly as possible. He told them that he would be hidden in the observation booth above and behind them but that they should, nevertheless, trust his judgment about when it would be safe for them to move between barriers and to open fire.

The leader told the participants that they would be running the course twice in the role of soldier, once with the buzzer active (touch condition) and once when it would not be active (the control condition). A series of coin tosses was used to determine in sequence in which the three participants played the roles of soldier and sniper.

The soldier's initial position was behind the barrier furthest from the sniper. After loading his or her rifle with the five paintballs from the small cup located behind the barrier, the soldier used the voice-activated head-set to inform the leader that he or she had reloaded ("laddat"). On the leader's order to fire, the solder attempted to knock down the four "hostile" (red) targets on the table in front of the sniper's bunker without hitting the "friendly" (blue) target. When out of ammunition, the soldier informed the leader that the rifle was empty ("slut"). On the leader's order to advance, the soldier moved to the next protective barrier on the

other side of the lane. The cycle was repeated at each of the eight barriers. The 40 paintballs, 5 at each barrier, were the soldier's only ammunition.

Dependent measures

The purpose of the targets was to give structure to the mission and the soldier something to do. In previous experiments we learned that measures of shooting accuracy revealed more about the participant's familiarity with firearms than about remote command and control. Accordingly, no data were gathered on when and how many targets the soldier knocked down.

The dependent variable was the time it took for the soldier to obey an order from the leader. In a series of pre-trials, we attempted to use two goniometers (strain gauges), one attached to leader's call-button finger and the other to the soldier's trigger finger, to obtain data on when orders were issued and when they were obeyed, respectively. This idea was scraped almost immediately when the goniometer attached to the second soldier in the test did not survive a high-velocity encounter with a paintball. We retreated to more reliable equipment, the hand-held stop watch. Any errors associated with fallible operation of the stopwatch are assumed to be randomly distributed within and across participants.

One of the experimenters was the timekeeper throughout the experiment. The timekeeper stood at one of the windows in the observation room where he could hear the leader issuing orders and see when it was completed. The criterion for completion differed across the two types of orders. For an order to fire, the time interval ended when the soldier first opened fire. For the order to move, the time interval ended with the soldier came to a full, controlled stop behind the next barrier. Procedures for coming to a controlled stop varied across participants. The within-subjects design and analysis controlled this inter-participant variability.

Results

This experiment was aimed at understanding whether experimental participants would obey orders more quickly when the orders were accompanied by a remote-controlled tactile stimulus applied between the shoulder blades. The within-subjects design had two conditions, the touch condition and a control condition, and two sequences of those conditions. In the touch condition, the soldier received a remote-controlled tactile stimulus and a verbal order at the same time. The control condition featured only the verbal order. The first trial for half the participants was the control condition; for the other half it was the touch condition. The hypothesis to be tested proposed that response times would be faster in the touch condition for both sequences of conditions. Accordingly, the hypothesis posits an interaction between condition and sequence and, when the data are collapsed across sequence, a main effect for the manipulation of touch.

Move times

Separate mixed-effects ANOVA were conducted for the response times to orders to move ("move times") and to fire ("fire times"). The means and standard errors for the move times are shown in Figure 3. The open columns in Figure 3 are the data for participants who ran the assault lane in the touch condition first. Accordingly the two shorter columns on the outside are data collected in the touch condition. As expected, response times were faster in the touch condition in both sequences of data acquisition. The ANOVA indicates the interaction is significant, $F(1, 44) = 14.08$, $p < .001$, $\eta^2 = .20$.

Because the effect of sequence is not significant, $F(1, 44) = 0.05$, $p > .8$, the data were combined and a within-subjects t test performed to gauge the effect of the manipulation of touch. The means and standard errors are shown in Figure 4. The manipulation was highly significant,

$t(23) = 6.75$, $p < .0001$, $\eta^2 = .66$. This result suggests that verbal orders to move across an assault lane where a sniper is maintaining live paintball fire are obeyed faster when accompanied by a tactile stimulus.

Fire times

The means and standard errors for the fire times are shown in Figure 5. Once again, the open columns are the data for participants who ran the assault lane in the touch condition first and the outer columns are data collected in the touch condition. Response times were virtually identical in all but the control-first trials. Participants who experienced the control condition first were markedly faster in the touch condition of second trial. In contrast, those who experienced the touch condition first did not slow down in the second trial. This pattern of data suggests they may be differential transfer of the prior experience of tactile stimulus. The ANOVA indicates that the interaction, $F(1, 44) = 3.40$, $p < .08$ is marginally significant as is the effect of sequence, $F(1, 44) = .29$, $p < .09$, $\eta^2 = .15$. Accordingly, the data were combined and a within-subjects t test performed to gauge the effect of the manipulation of touch. The means and standard errors are shown in Figure 6. The manipulation was again highly significant, $t(23) = 2.99$, $p < .004$, $\eta^2 = .28$. This result suggests that verbal orders to fire in the face of live paintball fire are obeyed faster when accompanied by a tactile stimulus.

The expected results obtain for both types of orders. Orders to move and to shoot issued by a remote leader were obeyed more quickly when accompanied by a tactile stimulus. The effect of the tactile stimulus was more pronounced for orders to move. The data strongly support the contention that the administration of a mild buzzing rumble between the shoulder blades had a strong impact on participants' performance under conditions emulating live fire and remote

command and control. The presence of marginally significant differential transfer, while unexpected, is auspiciously in the desired direction.

Qualitative data

No formal method for obtaining qualitative data was used. However, informal conversations were held with all participants after they completed both trials in the role of soldier. Many participants indicated that they did not notice the tactile stimuli at all. Some stated that they paid attention to the buzzing at first but then, rather quickly, ignored it entirely.

Post-hoc analyses

The experimenters observed that the effect of the tactile stimuli on move times appeared to increase as the soldiers approached the sniper's bunker, that is, as the risk of being shot increased. Indeed, as shown in Figure 7, the within-subject differences between the control and touch conditions in response times to orders to move tend to increase as the soldiers got closer to the bunker. The slope of the linear fit to the observed difference data, 0.10 second/move with a 95% confidence interval of .04 to .16, is significant, $t(159) = 3.45$, $p < .001$, supporting the casual observations. The trend is present only in the difference in times between the control and touch conditions and is not seen in the raw data from either condition alone. Thus, the apparent increase in the effectiveness of the tactile stimulus cannot be explained away as a function of increasing hesitancy in the control condition or of increasing disregard for personal safety in the touch condition. No trends of any kind were observed or found in analyses of the fire time data.

Discussion

The analyses indicate that the participants in the role of soldier were significantly faster to follow verbal orders when the command was accompanied by the tactile stimulus. These

results support the hypothesis that the administration of tactile stimuli can speed performance at obeying orders under conditions of remote command and control. The suggestion of differential transfer in the results for fire times is also encouraging. The possibility that prior experience with tactile stimuli may have lasting benefit lends further support to the claim that remote-controlled, simulated touch may become an effective tool for compensating, in part, for soldiers' hesitancy to obey remote leaders.

In spite of the participants' retrospective oblivion to the administration of simulated touch, they did obey orders more quickly when they received the tactile stimulus than when they did not. Accordingly, the phenomenon to be explained is not whether but how tactile stimuli act as an amplifying complement to a verbal order. It is entirely possible that the participants responded to the auditory information consciously and to the tactile stimuli subconsciously. These two ways of processing perceptual information are likely to be more or less complementary and neurologically parallel and independent (e.g., Damasio, 1999).

Whether or not we accept the argument that the tactile stimuli were processed without conscious awareness, the question of why touch promotes the intended goal-directed behavior remains. Two compatible alternatives suggest themselves: an executive control system (attention) and trust.

As shown in Figure 7, the difference in move times between the control and touch conditions tended to increase as the soldiers approached the sniper's bunker. The trend is found in the difference data but not in the raw response times. The implication is that the administration of a tactile stimulus led the soldiers to move relatively faster as they were exposed

to an ever-increasing risk of being shot by the sniper. The effect of remote-controlled touch appears to increase with repetition or as risk increases or both.

This observation suggests a behavioral explanation for the utility of tactile stimuli in remote command and control. The soldiers appeared to learn that the tactile stimuli were associated with times when it was safe to move. Accordingly, simulated touch may modulate the perception of risk through negative reinforcement; being shot is the unwanted stimulus that is removed. This interpretation suggests that the tactile stimuli reinforced the message that the remote leader's verbal commands could be trusted. Remote-controlled touch may spawn cognition-based trust in the remote leader which in turn leads to faster responses to orders.

Alternatively, the application of a tactile stimulus may have induced feelings, however unconscious, of leader presence and generated some form of affect-based trust in the remote leader. The induced trust, in turn, may have been sufficiently strong to lead the participants to disregard the risk to which they were exposed. However, this alternative is the less likely given the time required for affect-based trust to develop (McAllister, 1995).

It is interesting to note that the differential effect is found only in orders to move and not in orders to fire. The lack on an effect in differential fire times may reflect the fact that, when ordered to fire, the soldier's fire pins the sniper down and, accordingly, reduces the risk of being shot. Thus the exposure to risk while firing is relatively constant at all positions in the assault lane. The reduced risk of being shot while firing may obviate or lessen the need to depend on (to trust) the remote leader.

We interpret these results to show that the administration of tactile stimuli may have quickly reinforced the message that the remote leader was worthy of trust. This trust

compensated, in part, for the lack leader presence in this experimental paradigm. While the generality of this result remains to be tested, partial support can be found in the work of Milgram (1974) who found a connection between leader presence and participants' willingness to follow verbal orders.

It must be acknowledged that the benefit of the tactile stimuli may not be mediated by trust but solely by activation of a supervisory attentional system like that proposed by Norman and Shallice (1986). In this model of executive control, a supervisory attentional system becomes active and reinforces on-going goals when tasks are novel or not well-learned or are difficult or dangerous. Given that our paintball task was novel and arousing if not dangerous, it is reasonable to infer that responding to the remote leader's orders may set the stage for activation of an executive control system. Posner and DiGirolamo (1998) discuss neuroimaging data that show that executive control in verbal tasks is associated with selective activation of the anterior cingulate. The cingulate is also associated with processing of visceral, emotional, and haptic information (Damasio, 1999). Thus, the tactile stimuli may activate the anterior cingulate and its executive control system which, in turn, reactivates brain areas already primed by auditory input and quickens the motoric response.

There is no reason to believe that trust and executive control are mutually exclusive explanations for the observation of tactile reinforcement of verbal commands. It may well be the case that activation of executive control is the engine for the conditioning that develops cognitive-based trust in the remote leader.

If, as we suggest, trust is the intervening variable between simulated touch and the reduction in response times to verbal orders, much work remains to be done to assess whether its

benefits persist. There is a hint of stability in the differential transfer of the positive effect of the tactile stimulus on responses to orders to fire. An extensive series of trials would be needed to assess whether this benefit is stable or brittle and how it interacts with the outcomes of a leader's commands. It is likely that the military would want to design a reinforcement schedule for the administration of tactile stimuli in conjunction with verbal orders prior to adopting it as a routine part of remote command and control of fire teams.

An ergonomic issue is how best to apply the tactile stimuli so that they have the best possible effect. For example, it remains to be seen whether a tactile stimulus given on the trunk is the clearest experience of simulated touch and whether other stimuli may be more effective than a mild buzzing rumble at stimulating trust in a remote leader.

References

Adelson, M. (1961). Human decisions in command and control centers. *Annals of the New York Academy of Science, 89,* 726-731.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10,* 122-142.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81,* 211-241.

Brehmer, B. & Sundin, C. (2000). Command and control in network-centric warfare. In C. Sundin & H. Friman, (Eds.), *ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6* (pp. 45-56). Stockholm: The Swedish National Defense College.

Costa, A. C., Roe, R. A., & Taillieu, T. (2001). Trust within teams: The relation with performance effectiveness. *European Journal of Work and Organizational Psychology, 10*(3), 225-244.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace & Co.

DeSanctis, G. & Monge, P. (1999). Introduction to the special issue: Communication processes for virtual organizations. *Organization Science, 10*(6), 693-703.

Dirks, K. T. (2000). Trust in leadership and team performance: Evidence from NCAA basketball. *Journal of Applied Psychology, 85*(6), 1004-1012.

Eden, D. (1986). Team development: Quasi-experimental confirmation among combat companies. *Group and organization studies, 11*(3), 133-146.

Friman, H. (2000). Decision making in a ROLF staff environment. In C. Sundin & H. Friman, (Eds.), *ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6* (pp. 67-74). Stockholm: The Swedish National Defense College.

Guéguen, N. (2004). Nonverbal encouragement of participation in a course: The effect of touching. *Social Psychology of Education, 7*(1), 89-98.

Hall, J. (1996). Touch, status, and gender at professional meetings. *Journal of Nonverbal Behavior, 20,* 23-44.

Handy, C. (1995). Trust and the virtual organization. *Harvard Business Review, 73,* 40-50.

Hébert, R. (2003). Psychological science and the transformation of the military. *American Psychological Society Observer, 16*(9).

Henley, N. (1973). The politics of touch. In P. Brown (Ed.), *Radical psychology*. New-York: Harper & Row.

Hoffman, E., McCabe, K. A., & Smith, V. L. (1998). Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology. *Economic Inquiry, 36,* 335-352.

Jarvenpaa, S. L. & Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization Science, 10*(6), 791-815.

Johnson, S. & Levis, A. H. (1988). *Science of command and control: Coping with uncertainty*. Washington D. C.: AFCEA Press.

Jones, G. R. & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review, 23*(3), 531-546.

Kanawattanachai, P. & Yoo, Y. (2002). Dynamic nature of trust in virtual teams. *Journal of Strategic Information Systems, 11,* 187-213.

Kaempf, G. L., Klein, G., Throdsen, M., & Wolf, S. (1996). Decision making in complex command and control environments. *Human Factors, 38*(2), 220-231.

Lewicki, R, J. & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 114-139). Thousand Oaks, CA: Sage.

MacCrimmon, K. R. & Wehrung, D. A. (1985). A portfolio of risk measures. *Theory and Decision, 19*(1), 1-29.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 23*(3), 709-734.

McAllister, D. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*(1), 24-59.

Meyerson, D., Weick, K. E., & Kramer, R. M. (1996). Swift trust and temporary groups. In R. M. Kramer &, T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 166-195). Thousand Oaks, CA: Sage.

Milgram, J. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.

Norman, D. A., & Shallice, T. (1986). Attention to actions: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1-18). New York: Plenum Press.

Olson, G. & Olson, J. (1999). Distance matters. *Human-Computer Interaction, 15*, 139-178.

Orbell, J. M., Dawes, R. M., & van de Kragt, A. J. C. (1988). Explaining discussion–induced cooperation. *Journal of Personality and Social Psychology, 54*, 811-819

Pangburn, K., Freund, S., Pangburn, H., & Smith, K. (2003). Beyond Milgram: An experimental study of leader presence. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*, 364-368.

Posner, M. I. & DiGirolamo, G. J. (1998). Executive attention: Conflict, target detection, and cognitive control. In R. Parasuraman (Ed.), *The attentive brain* (pp. 401-423). Cambridge, MA: MIT Press.

Rocco, E. (1998). Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. *Proceedings of the Conference on Human Factors in Computing Systems - CHI'98*, 496-502.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *The Academy of Management Review, 23*(3), 393-404.

Smith, K. (in press). The adverse impact of remote command and control under live fire. *Human Performance in Extreme Environments*.

Smith, K. & Pangburn, K. (2004). *The value of leader presence*. Manuscript submitted for publication.

Throne, M. (2003). A military training mission. *American Psychological Society Observer, 16*(10).

Turner, J., Hogg, M., Turner, P., & Smith, P. (1984). Failure and defeat as determinants of group cohesiveness. *British Journal of Social Psychology ,23*, 97-111.

Wageman, R. (1995). Interdependence and group effectiveness. *Administrative Science Quarterly, 40*, 145-180.

Whitener, E. M., Brodt, S.E., Korsgaard, M.A., & Werner, J. M. (1995). Managers as initiators of trust: An exchange relationships framework for understanding managerial trustworthy behavior. *The Academy of Management Review, 23*(3), 513-530.

Yates, J. F. & Stone, E. R. (1992). The risk construct. In J. F. Yates (Ed.), *Risk-Taking Behavior* (pp. 1-25). New York: Wiley.

Appendix

Mission Brief

The Swedish version used in this experiment is available from the authors upon request.

Situation

The Krasnovian Army has taken an unknown number of allied POWs. These prisoners are being held in a firing line and are about to be executed. You are now an agent of a multinational force and have been assigned to rescue these prisoners.

Mission

Your mission is to sneak into the enemy camp and eliminate all enemy targets while minimizing friendly casualties without being hit by sniper fire.

Procedure

Under direction of your leader, move from position to position on the lane to advance toward the enemy camp and to secure ammunition. Hit hostile (red) targets without hitting the friendly (blue) target. To avoid sniper fire, your leader will inform you when hostile snipers are under cover and not firing.

1 When given the order “Fire” you will attempt to shoot the red targets from behind your current position.

2 When you are out of ammunition, report to your leader “Out of ammo.”

3 When it is safe to do so, your leader will direct you to move by giving you the order “Move.”

4 When you arrive at the next firing position, immediately pick up the container of ammunition and load your weapon. Five rounds of ammunition will be located at each position.

5 Once loaded, report to your leader by saying “Loaded.”

All orders will be given by your leader. Do not fire until told to. Do not move until told to. Is this understood?

Any questions?

Quiz

1 When do you fire your weapon?

2 What do you do when ordered to fire your weapon?

3 What do you do when you run out of ammunition?

4 When do you move to the next position?

5 What do you do when ordered to move?

6 What do you do immediately upon reaching the next position?

7 What do you do once you have reloaded your weapon?

Figure Captions

Figure 1. The layout of the paintball assault lane. The remote leader was positioned in an observation room above and behind the lane. Eight staggered barriers, four on each side, provided the participant playing the role of soldier protection from paintball fire from the participant playing the role of sniper. A cup containing five paintballs was located at each barrier, enabling the soldier to reload. Five targets (plastic bottles filled with colored water), four red (enemy forces) and one blue (a hostage to be rescued), sat on a shelf in front of the sniper's bunker.

Figure 2. The remote-controlled vibrator package. On the left is the tactile-stimulus unit, a shoulder belt with a small metal box containing a motor driving an eccentric cam. On the right is the communications unit and associated low-voltage power supplies. The stand-alone walkie-talkie was used by the remote leader to send a signal to initiate a tactile stimulus. The large box contained a sound-activated relay between the second, receiving walkie-talkie and the motor in the tactile-stimulus unit.

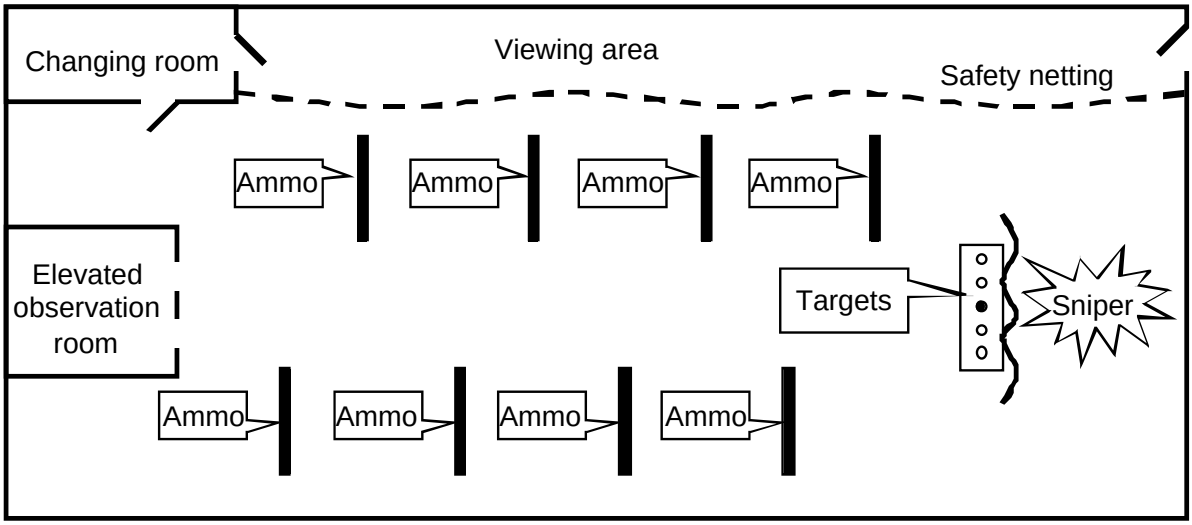
Figure 3. Mean and standard errors of response times to the order to move for the two sequences of data acquisition. The outer columns are data from the touch condition and the inner columns data from the control condition.

Figure 4. Mean and standard errors of response times to the order to move collapsed across trial sequence.

Figure 5. Mean and standard errors of response times to the order to fire for the two sequences of data acquisition. The outer columns are data from the touch condition and the inner columns data from the control condition.

Figure 6. Mean and standard errors of response times to the order to fire collapsed across trial sequence.

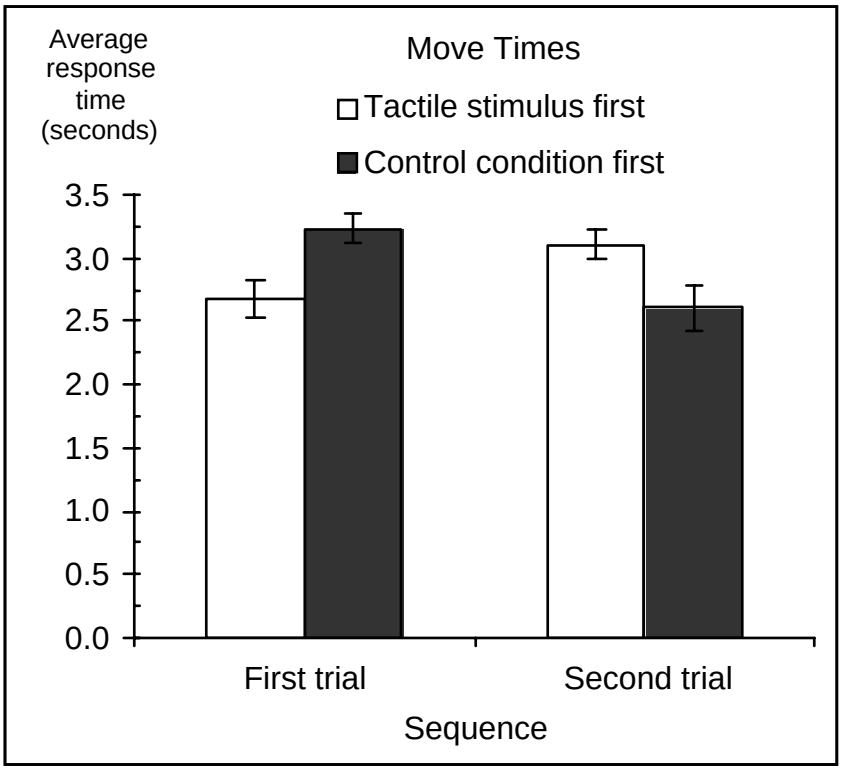
Figure 7. Observed differences in move times across conditions (control - touch) and fitted regression line showing an apparent linear increase in the benefit of the tactile stimulus as soldiers advanced toward the sniper's bunker.



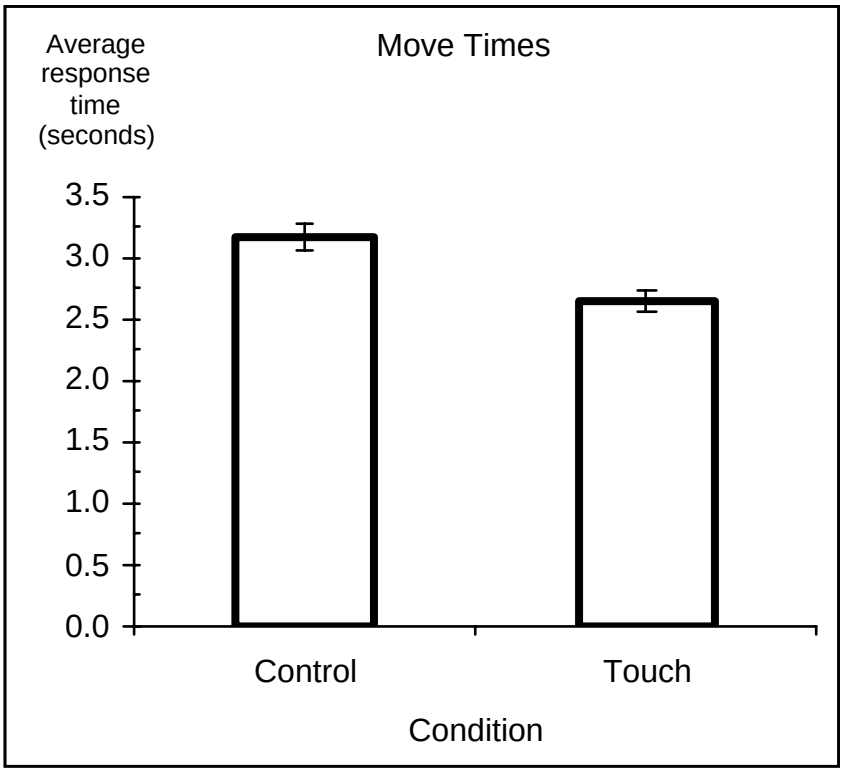
Simulated touch. Figure 1.



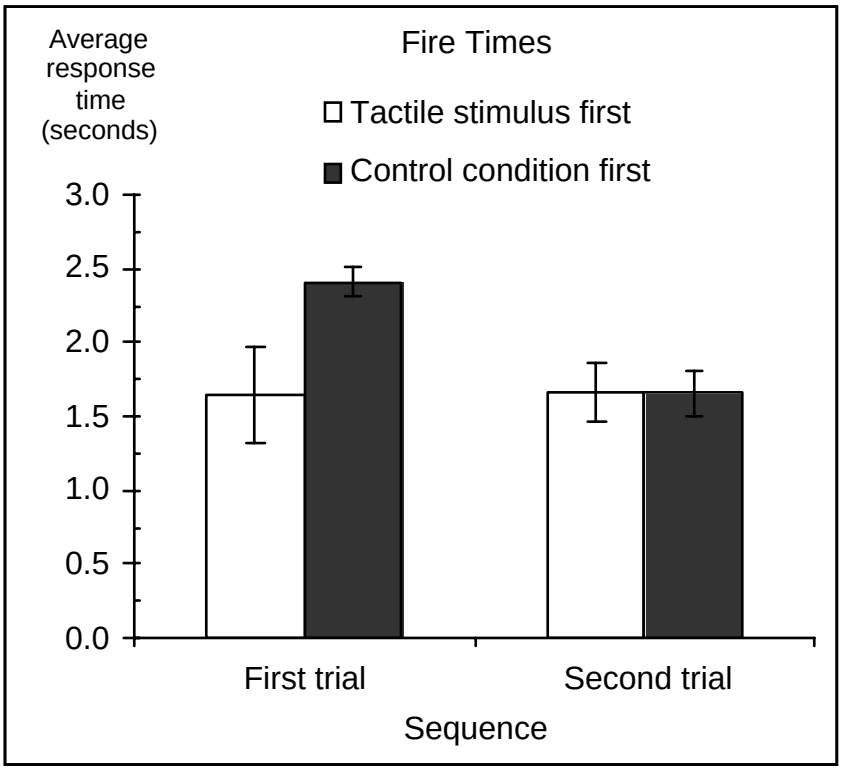
Simulated touch. Figure 2.



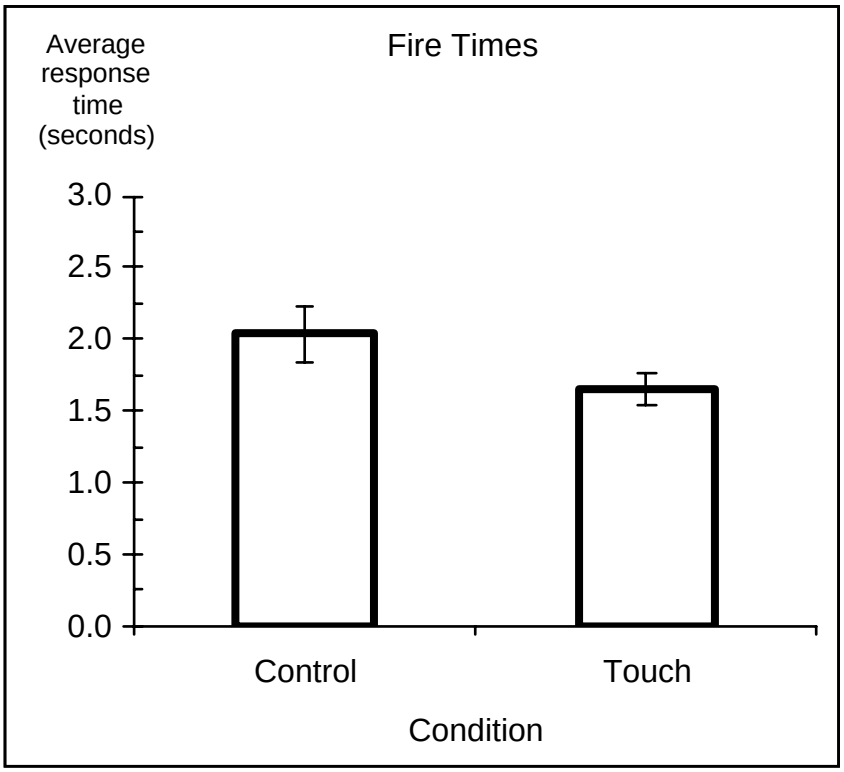
Simulated touch. Figure 3.



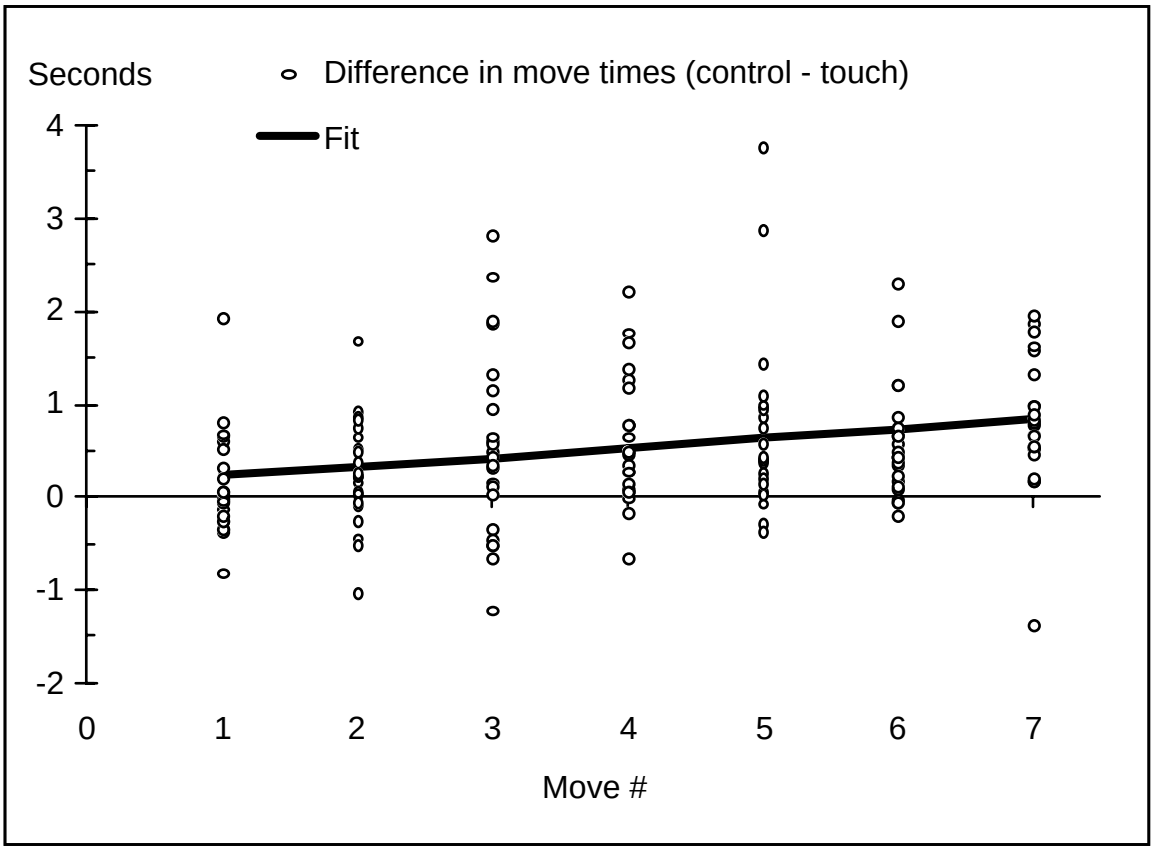
Simulated touch. Figure 4.



Simulated touch. Figure 5.



Simulated touch. Figure 6.



Simulated touch. Figure 7.

Running Head: LEADER PRESENCE

The Value of Leader Presence

Kip Smith

Division of Industrial Ergonomics

Linköping Institute of Technology

LT Keith Pangburn

74th QM Co., U. S. Army

Camp Anaconda, Iraq

Address correspondence to: Kip Smith, Division of Industrial Ergonomics, Linköping Institute of Technology, S-581 83 Linköping, Sweden, kipsm@ikp.liu.se, phone +46 13 28 27 64, fax +46 13 25 79

Abstract

Organizations face the *collocation dilemma* – should work teams and their leaders work in the same location and communicate face-to-face or form virtual teams that communicate via technologic links? At issue is the effect of organizational structure, collocated or distributed, on performance. For the military, the issue is whether soldier performance may be compromised by adopting the structure for distributed work known as remote command and control. To test the impact of remote command and control on soldier performance we have conducted a series of three experiments using live but non-lethal paintball fire. Participants playing the role of soldier obeyed verbal orders more quickly when collocated with a leader than when the leader was sheltered at a remote location. Communication mode had a negligible effect on performance. These findings support arguments against plans to adopt remote command and control of the dismounted infantry specifically and distributed work generally.

KEY WORDS: collocation, leader presence, remote command and control, mediated communication, paintball, fire team, trust, network based defense, objective force warrior

The Value of Leader Presence

One of the dilemmas facing organizations is whether work teams should be collocated or distributed (Iacono, 2002). Teams benefit from collocation, physical proximity, through direct observation, face-to-face conversation, nonverbal communication, and touch (Guéguen, 2004; Kiesler & Cummings, 2002). Collocation stimulates collaboration and the efficient, dynamic allocation of responsibilities (Clark & Brennan, 1990; Durso, Batsakes, Crutchfield, Braden & Manning, 2004). The personalized interactions afforded by collocation create social linkages that often develop into bonds that sustain team identity (Monge, Rothman, Eisenberg, Miller & Kirste, 1985; Nardi & Whittaker, 2002).

The forces pushing for distributed teams are both technologic and pragmatic. Technology has enabled the development of global organizations with operations in many countries and employees from diverse cultures (McGrath, Berdahl, & Arrow, 1995). Mediated communication (e.g., radio links, email, teleconferencing) frees the organization to allow employees to live where they prefer, to work autonomously, and to form 'virtual teams' (DeSanctis & Monge, 1999). The organization can assemble virtual teams on an as-needed basis to address specific short- or long-term goals, to take advantage of dispersed expertise, or to overcome culturally-based misunderstandings (Armstrong & Cole, 2002).

The United States and Swedish armed services have recently and independently addressed the collocation dilemma. The teams they are considering are called fire teams, the basic fighting unit of the dismounted infantry. The fire team consists of four to eight foot soldiers and, traditionally, a collocated officer. By tradition, fire teams and their officer have shared an arduous regimen of training. The mutual hardship builds camaraderie and leads them

to care for each others' welfare. In battle, the officer has typically fought side-by-side with the fire team, facing the same hazards as the soldiers. Orders and other communication have been mediated by speech, gesture, and physical contact. This collocation puts the officer in harm's way but cements the bonds of trust (McAllister, 1995; Meyerson, Weick, & Kramer, 1996). Every veteran of the dismounted infantry will attest to the significance of team identity forged, in part, by collocation.

For the military, the collocation dilemma is driven by both technologic and pragmatic concerns. Officers are a highly trained and valuable scarce resource. Putting them in harm's way makes little economic sense if a viable alternative can be found. Mediated communication, global positioning, and other technologies have the promise to provide the means for effecting the alternative known as *remote command and control*.

Remote command and control

Remote command and control is a structure for distributed work in patently hierarchical organizations like the military. The leader or a group of leaders (officers) sits out of harm's way in a control center. The work teams (fire teams) find themselves in geographically remote and potentially hostile locations. Remote command and control makes it relatively easy to change leaders as the situation facing the team evolves (Brehmer, 1992). This is critical to the military because modern urban warfare is highly dynamic and rarely predictable. After all, "Stuff happens" (D. Rumsfeld, April 12, 2003). The decisions made are often highly consequential and the actions, once initiated, are generally irrevocable. The environmental complexity and uncertainty of urban warfare and the criticality of the leader's decisions are often cited as the major desiderata for adopting remote, rather than collocated, command and control (Adelson,

1961; Brehmer & Sundlin, 2000). In addition, modern command and control centers are designed to give the leader(s) access to multiple sources of information that are not routinely available on the battlefield (Friman, 2000; Hébert, 2003). The centralization of decision making and of the flow of information facilitates the selective deployment of limited resources and speeds the cycle of decision making.

Research on remote command and control has a colorful tradition. In classic experiments on the 'obedience problem,' Stanley Milgram found that a leader's authority over a subordinate was a function of the leader's proximity (Milgram, 1974). The participants in Milgram's research were three times more likely to administer electric shocks to another person when instructed by a leader to do so when the leader was present than when the leader gave the order from a separate room via telephone. Milgram's finding that collocation is instrumental to obedience is highly relevant to the work discussed here since both are intended to generalize to the bond between an officer and a soldier in battlefield situations.

Leader presence and communication mode

One consequence of becoming a geographically distributed organization is that all communication between leaders and workers must be mediated by radios or other communications equipment. Accordingly, part of the proposals by the U. S. and Swedish armed services to adopt remote command and control of fire teams is to require all soldiers to carry various types of sensors and telemetry equipment so that the remote leader(s) can see and hear what the soldiers see and hear (Hébert, 2003).

Any transition from collocated to distributed work involves changes along two dimensions, leader presence and mode of communication. Collocated teams generally rely on

direct, face-to-face, verbal communication. Distributed teams are separated not only physically but also by the technology they use to communicate. Accordingly, there are completing hypotheses for any differences in performance that may be associated with a transition from collocated to distributed work. Differential performance may be attributable to leader presence or to communication mode or both.

Issues

While the inferred benefits of collocation has been documented in a wide range of relatively sedentary work environments (e.g., Kiesler & Cummings, 2002; Monge et al., 1985), the utility of leader presence to the efficient operation of a military command and control system has yet to be documented. To that end, we conducted a series of three experiments that manipulate leader presence and mode of communication in an ecologically realistic simulation of live fire.

The first two experiments contrast conditions of leader presence (collocated vs. remote) to address the issue of whether members of fire teams working with a collocated leader are faster to respond to commands to move and to shoot than members of fire teams working with a remote leader. Milgram's observations led us to expect faster response times to a collocated leader than to a remote leader. The third experiment introduces a third condition (collocated with a radio) to disambiguate the effects of leader presence and communication mode.

Experiment 1

Method

The dilemma we faced as experimenters was to create an experimental setting that simultaneously captured much of the look and feel of the battlefield and met the relevant standards for the ethical treatment of human subjects. The technology that made it palatable to run an experiment that exposed our participants to live fire is called paintball.

Paintball is a game that people of college age, mostly but not exclusively male, choose to play for fun. The source of fun is shooting non-lethal projectiles at your friends while trying to avoid being shot by them. Players (are required to) wear protective clothing and face masks. Concerns about allowing our participants shoot each other were disarmed by the facts that (1) paintball is a pastime for the participants' peer group, (2) we planned to provide all participants with protective gear, and (3) a registered nurse would be on site throughout the experiments. Thus, while the APA standards for the treatment of human subjects are mute with respect to live fire, we believe we were able to uphold the intent of those standards.

The purpose of the live paintball fire was to generate some anxiety so that the measures would more readily generalize to the battlefield. The major sources of stress were the fear of being shot and actually being hit by paintballs. The pain associated with being struck by a paintball is transient but real. Protective gear minimized the risk of injury. None of our participants received more than a few small bruises.

Participants

Twenty volunteers from Kansas State University (2 women, 18 men; median age 19, range 18 to 28) participated in Experiment 1. All were citizens of the United States. None had military or Reserve Office Training Corps (ROTC) experience.

Experimental setting and apparatus

Layout. The assault lane and its layout, shown in Figure 1, are based upon a training paradigm used extensively by the U. S. Army. The lane was delineated by eight staggered protective barriers, four on each side. A small cup containing five paintballs was placed behind each barrier. A protective wall containing two horizontal slots (0.5 m long by 0.2 m high) at shoulder height stood the far end of the lane. The wall formed the sniper's bunker. The slots allowed the sniper to shoot at (and be shot by) the soldier. Five targets, 1.5 liter plastic bottles filled with colored water, four red and one blue, were placed in random order in a row on a table directly in front of the slots in the sniper's bunker.

The US Army provided access to a heated 25 x 200 ft (9 x 61 m) building at Range 52, Fort Riley, Kansas. Fort Riley is the home of the US Army 1st Infantry and is an active training center for artillery. We set up our paintball assault lane in this building on several Saturdays.

Munitions. The paintball rifle was the armament used in the experiment. Ammunition consisted of varicolored paintballs. Participants, both the soldier and the sniper, were issued protective clothing consisting of a face shield, knee pads, elbow pads, and coveralls.

Design

Experiment 1 used two conditions of leader presence in a repeated measures design to test the effect of leader presence on the response times to verbal orders. In the collocated condition, the leader was present in the assault lane one barrier behind the participant playing the role of soldier. The collocated leader issued verbal orders to the soldier directly, that is, by yelling. In the remote condition, the leader stood outside the building where he could see but could not be seen by both the soldier and the sniper. By necessity, communication in the remote condition was mediated. The equipment used were walkie-talkies purchased at a local electronics supply store.

There were two types of orders, to move (advance to the next barrier in the lane) and to shoot (at the targets in front of the sniper's bunker) in both conditions. The sequence of conditions, collocated and remote, was randomly assigned across participants. An equal number of participants were assigned to the two sequences of conditions.

Tasks

Soldier. Every participant played two roles, soldier and sniper, in random order. The task when playing the role of soldier was to move forward from barrier to barrier and to fire at the targets upon receiving orders to do so. Each participant completed the mission two times in the role of soldier. As far as the soldier's mission was concerned, the collocated and remote trials were identical. The only difference was in how the orders were delivered and received.

Sniper. A second participant acted as the sniper. The sniper's task was to shoot the soldier, that is, to instill anxiety and to slow his or her advance up the assault lane. The opportunity to play the role of sniper was a major inducement to participation.

The sniper's bunker consisted of a protective wall at the end of the assault lane. The wall contained two slots that provided a clear view of the lane and insured that a solder who was not crouching behind a barrier would be exposed to the possibility of sniper fire. The sniper was directed by the leader in the presence of the soldier to pause briefly when moving from one of the slots in the wall to the other. The reason for directing the sniper to move slowly between the slots was to give the leader grounds for judging when it would relatively safe for the soldier to move and to shoot. The leader asked both the soldier and the sniper to explain the implications of these instructions before the first trial started. The sniper was then issued 40 paintballs and asked to take a position behind the protective wall.

Leader. The junior author, an army officer, played the role of leader in every session and condition throughout the experiment. The leader wore a standard battle-dress uniform and issued verbal orders to the participant playing the role of soldier. The actual words used to issue the orders were "move" and "fire." The timing of these orders was based upon the position of the sniper and the leader's judgment.

Procedure

Participants were met in groups of four to six at the entrance to the library on the campus of Kansas State University where they were given informed consent forms which they were asked to read and sign. The consent form advised them about the possibility of incurring some pain, the need to shoot paintballs at another person, and the fact that the research is supported by the Army Research Office of the United States of America. Both the driver and the consent form encouraged the participants to ask questions. Upon signing the consent forms, participants entered the university vehicle and were driven to Fort Riley. A one-way trip took 45 minutes

and required passing through a security gate and a variety of active firing range complexes. All told, the experiment took at least four hours of the participants' time.

During the ride to Fort Riley, participants were given and asked to read a paper copy of the mission brief reproduced as Appendix 1. The mission brief describes a fictitious scenario that provided the background for the task they would be asked to perform.

On arrival at the assault lane, participants were issued protective coveralls, a facemask, knee pads, and elbow pads and given instructions on the use and safety features of paintball rifles. They then met their (male) leader for the first time. The leader briefed participants using official military Operations Order format and addressed participants by their last names. The leader told them he would be giving them two types of orders, (1) to advance up the assault lane from one barrier to the next and (2) to shoot at the targets in front of the sniper's bunker. He instructed them to obey all orders as quickly as possible. The leader told the participants that they would be running the course twice in the role of soldier. A series of coin tosses was used to determine in sequence in which the participants played the roles of soldier and sniper.

Participants playing the role of soldier were sent down the lane individually. The soldier's initial position was behind the barrier furthest from the sniper. After loading his or her rifle with the five paintballs from the small cup located behind the barrier, the soldier informed the leader, either by yelling or by radio, that the rifle was "loaded." On the leader's order "fire," the solder attempted to knock down the four red (hostile) targets on the table in front of the sniper's bunker without hitting the blue (friendly) target. When out of ammunition, the soldier informed the leader that the rifle was "out of ammo." On the leader's order "move," the soldier advanced to the next protective barrier on the other side of the lane. The cycle was repeated at

each of the eight barriers. The 40 paintballs, 5 at each barrier, were the soldier's only ammunition. Whenever the participants took aim at the targets or moved between barriers, they risked exposure to the sniper's fire.

In the collocated condition, the leader remained one barrier behind the participant playing the role of soldier and communicated by yelling. In the remote condition, the leader went outside the building and observed the action through one of the many windows that lined both long sides of the building. The only contact between the leader and the soldier in the remote condition was by radio. The order of conditions was randomized across participants yielding two groups, a collocated-first group and a remote-first group.

Dependent measures

The purpose of the red and blue targets was to give structure to the mission and the soldier something to do. While developing the procedure in early versions of this study, we learned that measures of shooting accuracy revealed more about the participant's familiarity with firearms than about leader presence. Accordingly, no data were gathered on when and how many targets the soldier knocked down.

The dependent variable was the time it took for the soldier to obey an order from the leader as measured by a hand-held stop watch. Several research assistants served as timekeepers. The same researcher kept time in both trials for a given participant. Any errors associated with fallible operation of the stopwatch are assumed to be randomly distributed within and across participants.

The timekeeper was given a radio and hid (from the sniper's fire) where he or she could hear the leader issuing orders and see when the soldier complied. The criterion for completion

differed across the two types of orders. For an order to fire, the time interval ended when the soldier first opened fire. For the order to move, the time interval ended with the soldier came to a full, controlled stop behind the next barrier. Procedures for coming to a controlled stop varied across participants. The within-subjects design controlled this inter-participant variability.

Results

Move times

Figure 2a is a graph of the means and standard errors of response time to the command to move for both sequences of data acquisition. The open symbols show the data for the remote-first group. This group responded more slowly in the first trial when the leader was hidden outside the building than in the second trial when the leader was present in the lane. The closed circles, representing the collocated-first group, show that this group responded more quickly in the first trial, again when the leader was present in the lane.

A two-factor mixed-effects repeated-measures ANOVA was conducted to assess the significance of the manipulation of leader presence and the possibility of sequence effects. The sequence of data acquisition (collocated-first, remote-first) was the blocking variable at two levels. The repeated measure was the two conditions of leader presence. Neither sequence, $F(1, 18) = 0.120$, $p > .73$, nor the interaction of sequence with condition, $F(1, 18) = 0.407$, $p > .53$, was significant. As expected, leader presence was significant, $F(1, 18) = 8.476$, $p < .01$, $\eta^2 = 0.040$. This result suggests that the manipulation of leader presence influenced the speed with which participants responded to commands to move. Figure 2b shows the data aggregated across sequence to illustrate the significant difference in means across the manipulation of leader presence.

Fire times

Figure 3 presents the corresponding graphs for response times to the command to fire. The pattern of results is the same: both groups of participants responded more quickly when collocated with the leader. Once again, the mixed-effects repeated-measures ANOVA found that neither the sequence of data acquisition, $F(1, 18) = 0.781$, $p > .79$, nor the interaction of sequence and leader presence, $F(1, 18) = 0.778$, $p > .38$, was significant. As expected, the manipulation of leader presence was significant, $F(1, 18) = 5.307$, $p < .04$, $\eta^2 = 0.034$. These results suggest that leader presence influenced response times to commands to shoot. Figure 3b shows the aggregated data to illustrate the significant difference in means across the manipulation of leader presence.

Discussion

Drawing upon Milgram's work, we hypothesized that response times to verbal orders would be faster when the order was given by a collocated leader than by a remote leader. The data from Experiment 1 support this hypothesis.

The data are also consistent with an alternative hypothesis concerning the experimental procedure and setting. In the remote condition, the leader was outside the building and looked through its windows to determine when it would be relatively safe for participants playing the role of soldier to move and to shoot. The participants knew this. It is possible that the participants thought that the leader's external viewpoint deprived him of information that might be critically relevant to the timing of commands and, hence, to their own safety. By this hypothesis, the data showing an advantage for leader presence may merely reflect participants'

doubts about the completeness of information available to the remote leader occasioned by the experimental setting. Experiment 2 was designed to test this alternative hypothesis.

Experiment 2

To test the generality of the result from Experiment 1 and the alternative hypothesis concerning the completeness of information available to the remote leader, we moved the second experiment to an outdoors venue on-campus. The move had the secondary benefit of substantially reducing the travel time for participants.

Method

The method used in Experiment 2 largely replicated that used in Experiment 1. This section focuses on the differences in the setting and procedure between the experiments.

Experimental setting and apparatus

An assault lane identical to that used in Experiment 1 was set up adjacent to a complex of dormitories in a grassy meadow between two lines of mature elms. The outdoor setting was the only difference in the layout of the lane.

The absence of walls made the remote condition less remote. In Experiment 1, the wall of the building kept the remote leader completely out of sight. This relative invisibility may have led participants to doubt his ability to act in their best interest when issuing orders to move and to shoot. To assuage this source of doubt in Experiment 2, the remote leader hid behind a tree approximately 30 feet (9 m) behind the lane. Not only could the remote leader see and hear the lane without impediment, it was possible for the participant to turn around and, on occasion, see the remote leader.

Participants

The experiment was conducted on sunny Saturdays and invariably drew a crowd of students. Some, but far from all, of the onlookers were eager to participate. Participants came from this self-selected sample of convenience. The public forum served our purpose well by ensuring that our participants were fully informed of the nominal risks posed by live paintball fire. While self-selection violates the credo of random selection, it mirrors the process by which the United States conscripts young people for active duty. Conscripts are the population to whom we want these experiments to generalize.

Twenty-two students from Kansas State University (3 women,19 men; median age 19, range 18 to 26) participated in Experiment 2. As in Experiment 1, all were citizens of the United States and none had military or Reserve Office Training Corps (ROTC) experience. All three women and several of the men called themselves hunters and owned rifles.

Design, tasks, and dependent variables

As in Experiment 1, there were two conditions. In the collocated condition, the leader remained one barrier behind the participant playing the role of soldier and communicated by yelling. In the remote condition, the only contact between the leader and the soldier was by radio. The junior author once again played the role of leader throughout the experiment. Every participant played two roles, soldier and sniper, in random order. The task when playing the role of soldier was to move forward from barrier to barrier and to fire at the targets upon receiving orders to do so. Soldiers ran the lane twice, once in the collocated condition and once in the remote condition. The order of conditions was randomized. An equal number of participants were assigned to the remote-first and collocated-first groups. The dependent variable was the

time it took for the soldier to obey an order from the leader as measured by a hand-held stop watch. Several research assistants served as timekeepers. The same research assistant kept time in both trials for a given participant.

Procedure

Participants tended to volunteer in groups of two to four. They were met at a registration table by the nurse gave volunteers the informed consent form and the mission brief, answered their questions, instructed them on the use and safety features of paintball rifles, and scheduled their times for participation. The nurse, who never had to tend to a participant, was a valued member of the experimental team.

At the assigned times, participants were issued protective coveralls, a facemask, knee pads, and elbow pads. They then met their (male) leader for the first time. The leader briefed them on the tasks of soldier and sniper. A series of coin tosses was used to determine in sequence in which the participants played the roles of soldier and sniper.

Results

Move Times

The graph of Figure 4a shows the means and standard errors of response time to the command to move for both sequences of data acquisition in Experiment 2. The remote-first group, represented by the open circles, responded more quickly to commands to move in the second trial when the leader was present in the assault lane. It appears that the transition to a collocated leader, to getting yelled at face-to-face by a leader exposed to live fire, improved the performance of participants in the remote-first group. In contrast, the performance of

participants in the collocated-first group, shown by the solid circles, got no worse when taking orders from the remote leader.

The two-factor mixed-effects repeated-measures ANOVA found no effect of sequence (remote-first vs. collocated-first) on move time, $F(1, 20) = 2.19$, $p > .15$. The interaction of sequence with condition, $F(1, 20) = 4.322$, $p < .06$, was pragmatically significant. Thus, there may be a beneficial carry-over effect of previous experience with a collocated leader. As expected, the manipulation of leader presence, shown in Figure 4b, reveals a differential effect of leader presence, $F(1, 20) = 9.070$, $p < .01$, $\eta^2 = 0.132$.

Fire Times

Figure 5a shows the response time data for the command to shoot in Experiment 2. Open symbols represent the remote-first group and solid symbols the collocated-first group. Once again, both groups responded more quickly to commands to shoot when collocated with the leader in the assault lane. The ANOVA for fire times indicates no effect of sequence, $F(1, 20) = 2.676$, $p > .69$, no significant interaction of sequence with condition, $F(1, 20) = 0.907$, $p > .35$, and a significant effect of leader presence, $F(1, 20) = 5.259$, $p < .04$, $\eta^2 = 0.034$. As expected and illustrated in Figure 5b, leader presence made a significant difference in the speed with which participants obeyed commands to shoot.

Discussion

These results replicate those from Experiment 1 and support our original hypothesis that participants will be faster to obey verbal orders from a leader when collocated with the leader than when the leader is in a remote location. The difference in the degree of remoteness of the remote condition across the two experiments had a negligible effect on participant performance.

Accordingly, the data fail to support the alternative hypothesis that the experimental setting of Experiment 1 induced the delay in response times in the remote condition. The delay, or hesitation, is present even when there was no wall separating the remote leader from the participant playing the role of soldier. The hesitation occurred when participants were asked to expose themselves to live paintball fire under conditions simulating remote command and control.

Experiments 1 and 2 reveal that commands are obeyed more rapidly when issued by a collocated leader who yells than when issued by a remote leader who uses a radio. These two conditions leave unanswered the question of whether the observed hesitation is due to proximity or to the use of mediated communication. Experiment 3 introduces a third experimental condition that eliminates the confound of proximity and communication mode and directly tests these competing hypotheses.

Experiment 3

The first two experiments documented a consistent effect of leader presence. Participants responded more slowly to commands issued by a remote leader who used a radio (the remote condition) than to commands issued by a collocated leader who yelled (the collocated condition). Experiment 3 preserved these two conditions and added a third, the collocated-with-radio condition, to disambiguate the effects of leader presence and communication mode. In this new 'radio condition,' the leader was collocated in the assault lane and used a radio to communicate with the participant playing the role of soldier. If leader presence is the major source of the observed differences in response times seen in the first two experiments, then performance in the remote condition should be slower than in both the yell and radio conditions. Further,

performance in the yell and radio conditions should be approximately the same. In contrast, if the differences are due to mediated communication, then performance in the yell condition should be faster than in both the remote and radio conditions and performance in the radio condition should be similar to that in the remote condition.

Note that it makes no sense to stage the complementary fourth condition (a remote leader who yells). This fundamental constraint on distributed work precludes a fully crossed design with two levels of leader presence and two levels of communication mode.

Method

The method used in Experiment 3 largely replicated that used in Experiment 1. This section focuses on the differences in the setting and procedure between the experiments.

Participants

Eighteen volunteers participated in Experiment 3 (7 women, 11 men; median age 24, range 20 to 31). Most were students of nursing or were associated with the Department of Cognitive Science at the Högskolan (college) in Skövde, Sweden.

This sample differs in several respects from those in the first two experiments. The participants were all Scandinavian (15 Swedes, 2 Icelanders, and 1 Norwegian) and, as is typical of college students in Scandinavia, were older and noticeably more mature than their American counterparts. Six had military experience of either 10 or 24 months. Of the Swedes, all but four had hunted game and owned rifles.

In spite of these cultural and age differences, there is little reason to expect to these factors to induce a significant difference in the willingness to obey commands from an

unfamiliar leader with a military bearing. If a hypothesis were to be advanced based on cultural differences, it would likely claim that the Scandinavians, being from a culture that values interdependence (e.g., Hofstede, 1980; Markus & Kitayama, 1991), would respond more willingly and, perhaps, more quickly than their more independently-minded American counterparts. Similarly, it would be reasonable to suggest that prior military experience should hasten compliance. While such speculation is intriguing and has multinational application (Sutton, 2003), testing hypotheses about cross-cultural differences is beyond the scope of this study and cannot be supported given the relatively small sample size.

Experimental setting and apparatus

Layout. The experiment was conducted indoors at a commercial paintball facility in Tidan, near Skövde, Sweden. The layout of the assault lane was essentially identical to that shown in Figure 1 but the dimensions of the lane were slightly larger and the barriers far more robust. The assault lane was delineated by eight large (2.5 m long by 1.5 m high) inflatable protective barriers, four on each side, 8 m apart. As before, a small cup containing five paintballs was placed behind each barrier. A protective wall containing two horizontal slots (0.5 m long by 0.2 m high) at shoulder height stood at the far end of the lane and formed the sniper's bunker. A table made of shipping pallets was placed directly in front of the slots in the sniper's bunker. Five targets, 1.5 liter plastic bottles filled with colored water, four red and one blue, were placed in random order in a row on the table.

An observation room above and behind the assault lane contained two screened windows that provided a commanding view of the lane, the soldier, and the sniper below. The time-keeper

and the leader in the remote condition stood in this room away from the windows to avoid being inadvertently seen by the participant playing the role of soldier.

Munitions and communications. The paintball rifle was the armament used in the experiment. Ammunition consisted of either green or yellow paintballs. Participants, both the soldier and the sniper, were issued protective clothing consisting of a face shield and coveralls. The communication equipment used to issue verbal orders was Danita SP3380 walkie-talkies equipped with and voice-activated by headsets.

Design

This experiment used three conditions of leader presence in a repeated-measures design. In the remote condition, the leader issued orders from the observation room above and behind the lane using a radio. In the yell condition, the leader was present in the assault lane one barrier behind the soldier and communicated with the soldier by yelling. The remote and yell conditions replicated the remote and collocated conditions used in Experiments 1 and 2, respectively. In the radio condition, the leader was present in the assault lane one barrier behind the soldier and used a radio to issue orders. Three of the six possible sequences of the three conditions (Yell first, Radio, Remote; Radio first, Yell, Remote; and Remote first, Yell, Radio) were used to collect data. Six participants were randomly assigned to each sequence. As in the previous experiments, there were two types of orders, to move (advance to the next barrier in the lane) and to shoot (at the targets in front of the sniper's bunker).

Tasks

Soldier. Every participant played two roles, soldier and sniper, in random order. Each participant completed the mission three times in the role of soldier. In the radio and remote

conditions, the soldier used a radio equipped with a headset to communicate verbally with the leader.

Sniper. A second participant acted as the sniper. The sniper was issued 40 paintballs. The sniper's task was to shoot the soldier, that is, to instill anxiety in the soldier and to slow his or her advance up the assault lane. On average, the Scandinavian snipers appeared to be much more judicious and effective in their fire than their American counterparts.

Leader. Three different officers from the military garrison in Skövde played the role of leader. These officers, all male, were dressed in military fatigues and had a military bearing. The leader issued verbal orders to the participant playing the role of soldier in Swedish. The orders that were given in all three conditions were "eld" ("fire") and "framåt" ("advance").

Procedure

Participants were picked up in groups of three at the University of Skövde and given an informed consent form which they were asked to read and sign. The consent form advised them, in Swedish, about the possibility of incurring some pain, the need to shoot paintballs at another person, the fact that the research is supported by the Army Research Office of the United States of America, and encouraged them to ask questions. They were then taken by car to the indoor paintball arena in Tidan. During the twenty-minute ride, they were given and asked to read a paper copy of the Swedish translation of the mission brief included in the appendix.

On arrival in the paintball arena, the participants were issued protective coveralls and a facemask. A member of the club that runs the arena instructed them on the use and safety features of paintball rifles. Participants then met their leader for the first time. The leader told the participants in Swedish that he would be giving them two types of orders and to obey all

orders as quickly as possible. The leader briefed the group of participants at the same time about the roles of soldier and sniper and invoked the same stipulations on sniper behavior as in the first two experiments. A series of coin tosses was used to determine the sequence in which the participants played the roles of soldier and sniper.

The soldier's initial position was behind the barrier furthest from the sniper. After loading his or her rifle with the five paintballs from the small cup located behind the barrier, the soldier either yelled or used the voice-activated head-set to inform the leader that he or she had reloaded ("laddat"). On the leader's order to fire, the solder attempted to knock down the four "hostile" (red) targets on the table in front of the sniper's bunker without hitting the "friendly" (blue) target. When out of ammunition, the soldier informed the leader that the rifle was empty ("slut"). On the leader's order to advance, the soldier moved to the next protective barrier on the other side of the lane. The cycle was repeated at each of the eight barriers. The 40 paintballs, 5 at each barrier, were the soldier's only ammunition.

Dependent measures

The dependent variable was the time it took for the soldier to obey an order from the leader. In a series of pre-trials, we attempted to use two goniometers (strain gauges), one attached to the leader's right index finger and the other to the soldier's right index (trigger) finger, to obtain data on when orders were issued and when they were obeyed, respectively. This idea was scrapped almost immediately when the goniometer attached to the second soldier in the test did not survive a high-velocity encounter with a paintball. We retreated to more reliable equipment, the hand-held stop watch. The lead author was the timekeeper throughout the experiment. The timekeeper stood at one of the windows in the observation room where he

could hear the leader issuing orders and see when it was completed. Any errors associated with fallible operation of the stopwatch are assumed to be randomly distributed within and across participants.

Results

Two-factor mixed-effects repeated-measures ANOVA were calculated for both the move time and fire time data. The blocking variable was the three sequences of data acquisition (e.g., Yell first, Radio, Remote); the repeated measure was the three conditions of leader presence.

Move times

Figure 6 is a graph showing the mean and standard errors for response times to the command to move for the three conditions of leader presence. Participants took appreciably longer to respond to the leader's command to move in the remote condition. The ANOVA revealed no effect of sequence of data acquisition, $F(2, 15) = .0605$, $p > .55$, and no interaction of sequence and condition, $F(4, 30) = 0.232$, $p > .18$. As expected, the main effect for leader presence was significant, $F(2, 30) = 4.370$, $p < .03$, $\eta^2 = 0.047$. The Tukey HSD post-hoc comparison of means indicated that the remote condition differed from radio condition at alpha = 0.05 and from the yell condition at alpha = 0.10.

Fire times

Figure 7 is the corresponding graph for fire times. Once again, participants took appreciably longer to respond to the leader's command to fire in the remote condition. The ANOVA revealed no effect for the sequence of data acquisition, $F(2, 15) = 0.605$, $p > .55$, and no interaction of sequence with condition, $F(4, 30) = 0.725$, $p > .58$. The main effect for leader

presence was significant, $F(2, 30) = 3.353$, $p < .05$, $\eta^2 = 0.058$. The Tukey HSD post-hoc comparison of means indicated that the remote condition differed from both the radio and yell conditions at alpha = 0.10.

Discussion

Additional participants were not recruited once three criteria were met: an equal number had participated in each block, the results replicated those in the first two experiments, and all measures of interest achieved statistical significance. As in the first two experiments, participants playing the role of soldier responded more slowly to a remote leader than to a collocated leader. This result has now been shown to obtain across cultures and in a slightly older population.

The addition of the radio condition disambiguates the effects of leader presence and mediated communication. Response times in the radio condition were statistically indistinguishable from those in the yell condition. Using radios did not influence the participants' willingness to obey a collocated leader's orders. It appears that the observed hesitancy in the remote condition is due to the lack of leader presence and not to mediated communication.

General Discussion

Remote command and control and soldier performance

The armed forces of the United States and Sweden are planning to change the organization and training of the dismounted infantry in a fundamental way. One of the cornerstones of their independently-derived but similar plans is a reliance on remote command

and control. With the advent of the 'Future force' or 'Objective force warrior' in the U. S. Army and 'Network-based defense' in the Swedish Army, fire teams may no longer receive orders in battle from a leader within visual range. Instead, their only connection with their commanding officers may be their radios and other portable information devices. Further, the fire teams may be assembled on an as-needed basis and may never have trained together or with their leader. Our experiments replicate an extreme realization of the proposed organizational change: participants did not know the leader and had insufficient time to become familiar with the leader and to develop mutual trust (e.g., McAllister, 1995; Meyerson, Weick, & Kramer, 1996).

In our experiments, the mode of communication with an unfamiliar leader, by yelling or by radio, had no effect on the speed with which participants followed that leader's orders. However, participants playing the role of soldier were consistently slower to expose themselves to live paintball fire when the leader was out of harm's way than when the leader was exposed to the same live fire. This result does not bode well for the adoption of remote command and control of fire teams led by leaders who have not had the opportunity to meet and become known by their soldiers.

The leaders in our experiments were professional military officers. Their bearing and the associated reputation effects appear to have been insufficient to overcome our participants' hesitation to obey commands issued by an unfamiliar remote leader.

These results raise several theoretical and empirical questions. Pragmatically, it will be important for the army to know whether active-duty soldiers exhibit the same hesitancy to remote command and control as our participants and whether familiarity with the leader eliminates the effect. A series of experiments with active duty soldiers is planned to address

these issues. A more far-reaching question concerns why participants hesitated to obey a remote leader. Our working hypothesis focuses on trust.

Trust in the leader

Trust is the willingness of one party to be vulnerable to the actions of another based upon positive expectations of the other's behavior (Rousseau, Sitkin, Burt, and Camerer, 1998). The critical elements for its formation are reciprocal vulnerability and interdependence (Berg, Dickhaut, & McCabe, 1995; Hoffman, McCabe, & Smith, 1998; McAllister, 1995; Meyerson, Weick, & Kramer, 1996). Where there is no interdependence and no vulnerability, levels of trust are likely to be low (Jones & George, 1998; Klein, Feltovich, Bradshaw, & Woods, in press).

In the collocated (yell and radio) conditions, the leader was vulnerable to paintball fire. The collocated leader assumed the same risk as the participant playing the role of soldier. The leader was unarmed and had to rely on the soldier's fire to suppress the sniper's fire. The collocated leader and soldiers were reciprocally vulnerable and interdependent. A tacit awareness of these factors may have generated sufficient trust in the collocated leader to promote faster responses to the leader's commands.

Retreating out of harm's way in the remote condition removed the foundation for the formation or maintenance of trust. It appears that remote command and control by an unfamiliar leader may preclude or substantially undermine the development of trust. These considerations lead us to believe that trust in the leader is the mediating variable that explains the hesitation observed in the remote condition.

Generality

The consistency of the results across venues and cultures lends credence to the hypothesis that leader presence is preferable to remote command and control in situations like those encountered by fire teams. Three of the factors that shape such situations are a highly dynamic and uncontrollable environment, the demand for a rapid response, and the imperative to adhere to authority (Brehmer & Sundin, 2000). To varying degrees, these factors shape the work of distributed teams in non-military settings as well. While the results reported here generalize most directly to fire teams, their implications extend to many organizations that operate in 'virtual' environments.

The small sample sizes used in these experiments preclude drawing any inferences about the influence of culture, gender, or military experience on willingness to obey a leader. Of these three factors, only military experience is directly relevant to the hypothesis being tested. The question remains whether active duty soldiers would display the same hesitancy under remote command and control as our civilian participants. We have planned a longitudinal experiment with Swedish recruits as participants to investigate the influence of basic training on the effect of leader presence on response times to verbal orders.

The experiments reported here did not use a double-blind procedure. They are, accordingly, susceptible to criticisms regarding the possibility of experimenter influence and demand effects. In the planned series of experiments, active duty soldiers and officers who are blind to the hypotheses will serve as timekeepers and leaders.

Mitigation

The results of these experiments compound the collocation dilemma for the United States and Swedish armed services. On the one hand, the experiments suggest that leader presence may be essential to soldier efficiency. On the other, military leadership is committed to effecting a smooth transition to remote command and control. To resolve this dilemma it will be necessary either to eliminate or to compensate for the observed hesitation to commands issued by a remote leader (or both). Two approaches suggest themselves, training and technologic intervention.

It is possible that training aimed at developing 'swift trust' in the leader will largely compensate for the inherent disadvantage of a distributed organizational structure. Given the ad-hoc, on-demand composition of many distributed teams, such training would have to meet strict demands for flexibility and effectiveness.

An alternative would be the development of technology that enables 'virtual leaders' to take the field with their (fire) teams. The requirements for a virtual leader are not physical or holographic presence but psychological presence. We plan to test alternative designs for information telemetry and display that aim to offset the decrements in performance that accompany remote command and control.

The paintball paradigm

These experiments demonstrate the viability of the paintball arena as an experimental setting for studying behavior under live fire and the special case of distributed work known as remote command and control. The unfolding of events in the paintball arena occupy a middle position between the strict controls of the experimental laboratory and the uncontrollable dynamics of the natural world. As such, the paintball arena lends itself as a paradigm for

conducting rigorous studies of both individuals and teams responding to realistic and challenging sources of stress. The potential applications are not limited to leader presence. Commercial facilities market paintball as an intervention that can promote team cohesion.

The downside of conducting experiments using the paintball paradigm is their relatively high cost. The experimenter must either invest in expensive equipment with a dubious record of reliability (jamming) or rent equipment, space, and technical know-how from a commercial facility. Our experience with both options leads us to recommend the latter. Renting increases the monetary expense but eliminates numerous sources of inconvenience and delay. It also relieves the experimenter of the responsibility of cleaning up the inevitable mess.

Acknowledgements

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This paper has benefited from presentations before the Human Factors and Ergonomics Society, the Society for Human Performance in Extreme Environments, and the Second Conference on Human Performance, Situation Awareness and Automation Technology and from comments by Laurel Allender, Peter Hancock, Donald Lampton, Robert Pleban, and Elmar Schmeisser. This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621.

We thank the many people who have helped make these experiments possible. A partial list includes Lee Bridenstein, Monica Clement, Paul Hemeren, Jens Hilmersson, Magnus Jacobsson, Andy Kearney, Arthur DeGroat, Martin Liberg, Jessica McGinnis, Holly Pangburn, Diane Ungvarsky, the owner and staff of Krazy Kris' Paintball Range, Keats, Kansas, and the members of Spelapaintball, Tidan, Västergötland, Sverige.

References

Adelson, M. (1961). Human decisions in command and control centers. *Annals of the New York Academy of Science, 89,* 726-731.

Armstrong, D. L. & Cole, P. (2002). Managing distances and differences in geographically distributed work groups. In P. Hinds and S. Kiesler (eds.), *Distributed Work* (pp. 167-186). Cambridge, MA: MIT Press.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10,* 122-142.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81,* 211-241.

Brehmer, B. & Sundin, C. (2000). Command and control in network-centric warfare. In C. Sundin & H. Friman, (Eds.), *ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6* (pp. 45-56). Stockholm: The Swedish National Defense College.

Clark & Brennan, (1990). Grounding in communication. In L. Resnick, J. Levine, and S. Teasley (eds.), *Perspectives on socially shared cognitio*n (pp. 127-149). Washington, DC: American Psychological Association Press.

DeSanctis, G. & Monge, P. (1999). Introduction to the special issue: Communication processes for virtual organizations. *Organization Science, 10*(6), 693-703.

Durso, F. T., Batsakes, P. J., Crutchfield, J. M., Braden, J. B. & Manning, C. A. (2004). The use of flight progress strips while working live traffic: Frequencies, importance, and perceived benefits. *Human Factors, 46*(1), 32-49.

Friman, H. (2000). Decision making in a ROLF staff environment. In C. Sundin & H. Friman, (Eds.), *ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6* (pp. 67-74). Stockholm: The Swedish National Defense College.

Guéguen, N. (2004). Nonverbal encouragement of participation in a course: The effect of touching. *Social Psychology of Education, 7*(1), 89-98.

Hébert, R. (2003). Psychological science and the transformation of the military. *American Psychological Society Observer, 16*(9).

Hoffman, E., McCabe, K. A., & Smith, V. L. (1998). Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology. *Economic Inquiry, 36*, 335-352.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* Newbury Park, CA: Sage Publications.

Iacono, C. S. (2002). Preface. In P. Hinds and S. Kiesler (eds.), *Distributed Work* (pp. xiii-xviii). Cambridge, MA: MIT Press.

Jones, G. R. & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review, 23*(3), 531-546.

Kiesler, S. & Cummings, J. N. (2002). What do we know about proximity and distance in Work groups? A legacy of research. In P. Hinds and S. Kiesler (eds.), *Distributed Work*. (167-186). Cambridge, MA: MIT Press.

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (in press). Common ground and coordination in joint activity. In W.R. Rouse and K. B. Boff (Eds.) *Organizational simulation*. John Wiley and Sons.

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224-253.

McAllister, D. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*(1), 24-59.

McGrath, J. E., Berdahl, J. L., & Arrow, H. (1995). Traits, expectations, culture, and clout: The dynamics of diversity in work groups. In S. E. Jackson & M. N. Ruderman (eds.)., *Diversity in work teams: Research paradigms for a changing workplace* (pp. 17-45). Washington, DC: American Psychological Association Press.

Meyerson, D., Weick, K. E., & Kramer, R. M. (1996). Swift trust and temporary groups. In R. M. Kramer &, T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 166-195). Thousand Oaks, CA: Sage.

Milgram, S. (1974). *Obedience to Authority*. New York: Harper & Row.

Monge, P., Rothman, L., Eisenberg, E., Miller, K., & Kirste, K. (1985). The dynamics of organizational proximity. *Management Science, 31*(9), 1129-1141.

Nardi, B. A., & Whittaker, S. (2002). The place of fact-to-face communication in distributed work. In P. Hinds and S. Kiesler (eds.), *Distributed Work* (pp. 83-110). Cambridge, MA: MIT Press.

Pangburn, K., Freund, S., Pangburn, H., & Smith, K. (2003). Beyond Milgram: An experimental study of leader presence. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*, 364-368.

Smith, K. (in press). The adverse impact of remote command and control under live fire. *Human Performance in Extreme Environments*.

Sutton, J. L. (2003). Validation of cultural awareness training concept. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*.

Appendix

Mission Brief

The Swedish version used in this experiment is available from the authors upon request.

Situation

The Krasnovian Army has taken an unknown number of allied POWs. These prisoners are being held in a firing line and are about to be executed. You are now an agent of a multinational force and have been assigned to rescue these prisoners.

Mission

Your mission is to sneak into the enemy camp and eliminate all enemy targets while minimizing friendly casualties without being hit by sniper fire.

Procedure

Under direction of your leader, move from position to position on the lane to advance toward the enemy camp and to secure ammunition. Hit hostile (red) targets without hitting the friendly (blue) target. To avoid sniper fire, your leader will inform you when hostile snipers are under cover and not firing.

1 When given the order “Fire” you will attempt to shoot the red targets from behind your current position.

2 When you are out of ammunition, report to your leader “Out of ammo.”

3 When it is safe to do so, your leader will direct you to move by giving you the order "Move."

4 When you arrive at the next firing position, immediately pick up the container of ammunition and load your weapon. Five rounds of ammunition will be located at each position.

5 Once loaded, report to your leader by saying "Loaded."

All orders will be given by your leader. Do not fire until told to. Do not move until told to. Is this understood?

Any questions?

Quiz:

1 When do you fire your weapon?

2 What do you do when ordered to fire your weapon?

3 What do you do when you run out of ammunition?

4 When do you move to the next position?

5 What do you do when ordered to move?

6 What do you do immediately upon reaching the next position?

7 What do you do once you have reloaded your weapon?

Figure captions

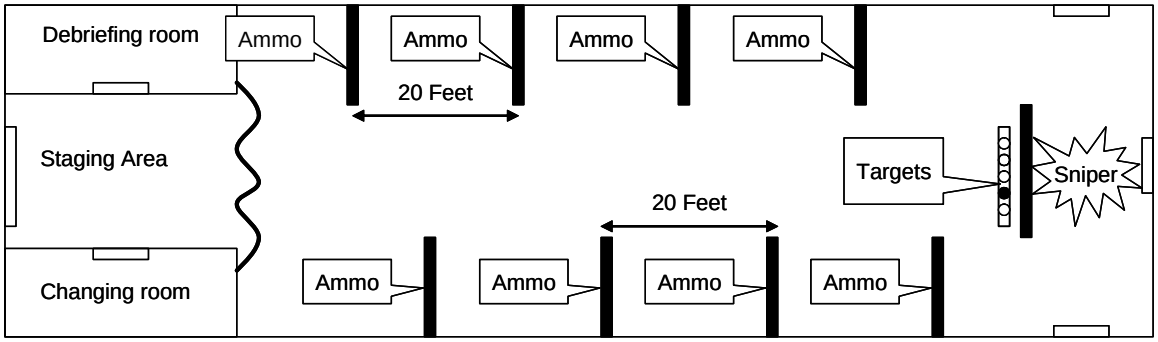


Figure 1 The layout of the paintball assault lane for Experiment 1, conducted inside a 200 ft x 30 ft (61 m x 9 m) Quonset hut at Fort Riley, Kansas. The layout was modified slightly for Experiment 2 to accommodate its outdoor setting in a tree-lined glade on the campus of Kansas State University. The layout for Experiment 3, conducted at the indoor training arena of Spelapaintball in Tidan, Sweden, was essentially identical to that shown here.

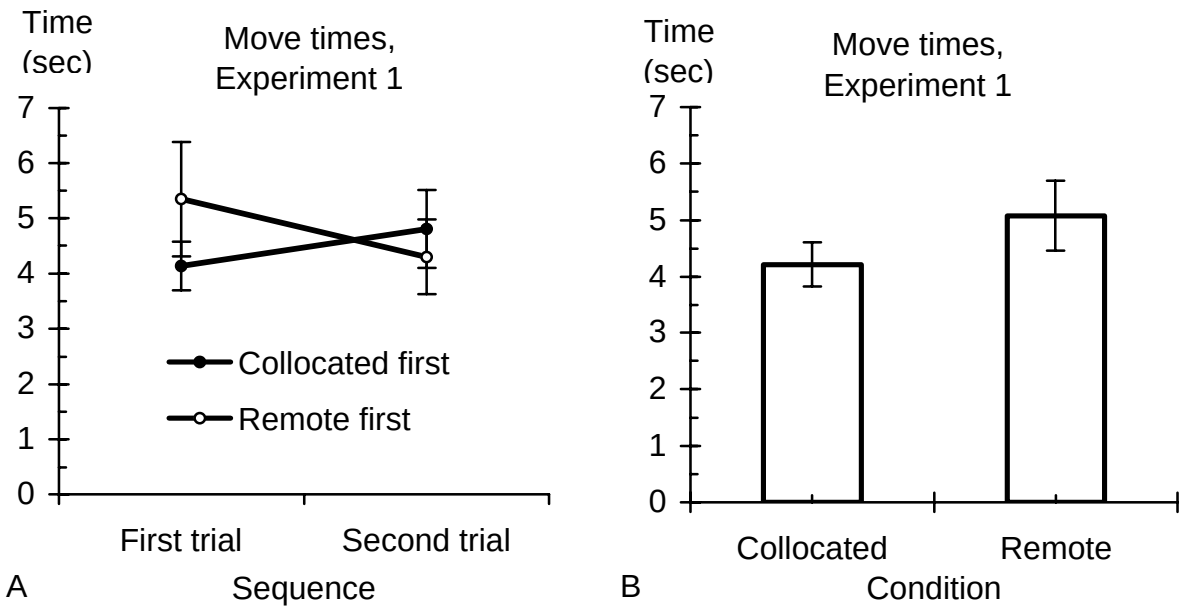


Figure 2 Average response times to commands to move (A) and the distribution of differences in response times between the remote and collocated conditions (B), Experiment 1.

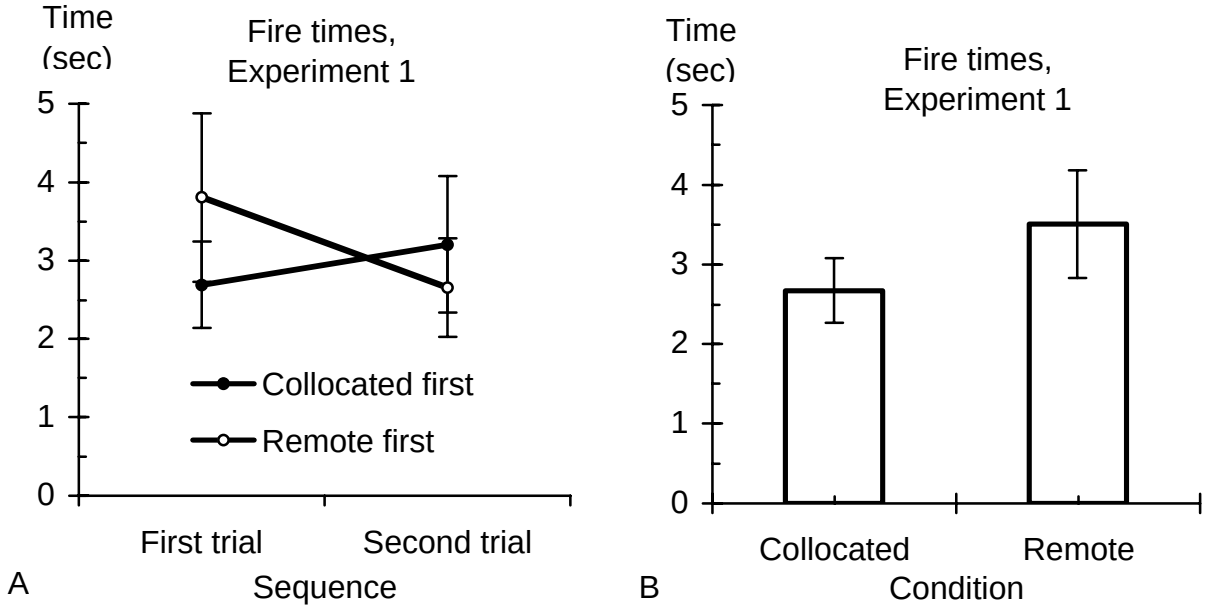


Figure 3 Average response times to commands to fire (A) and the distribution of differences in response times between the remote and collocated conditions (B), Experiment 1.

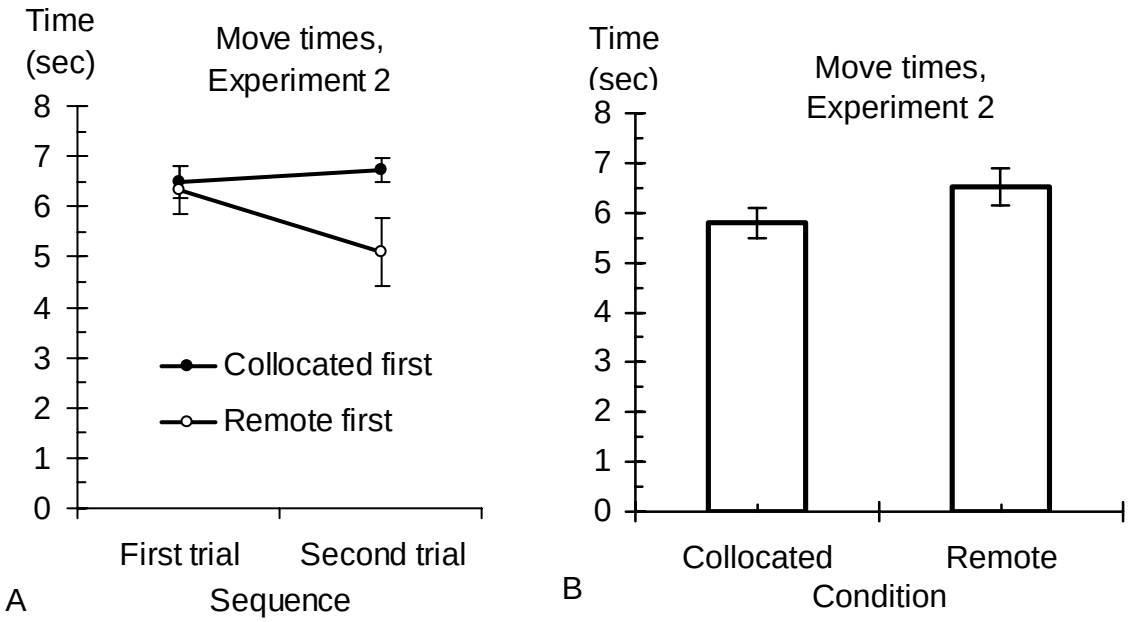


Figure 4 Average response times to commands to move (A) and the distribution of differences in response times between the remote and collocated conditions (B), Experiment 2.

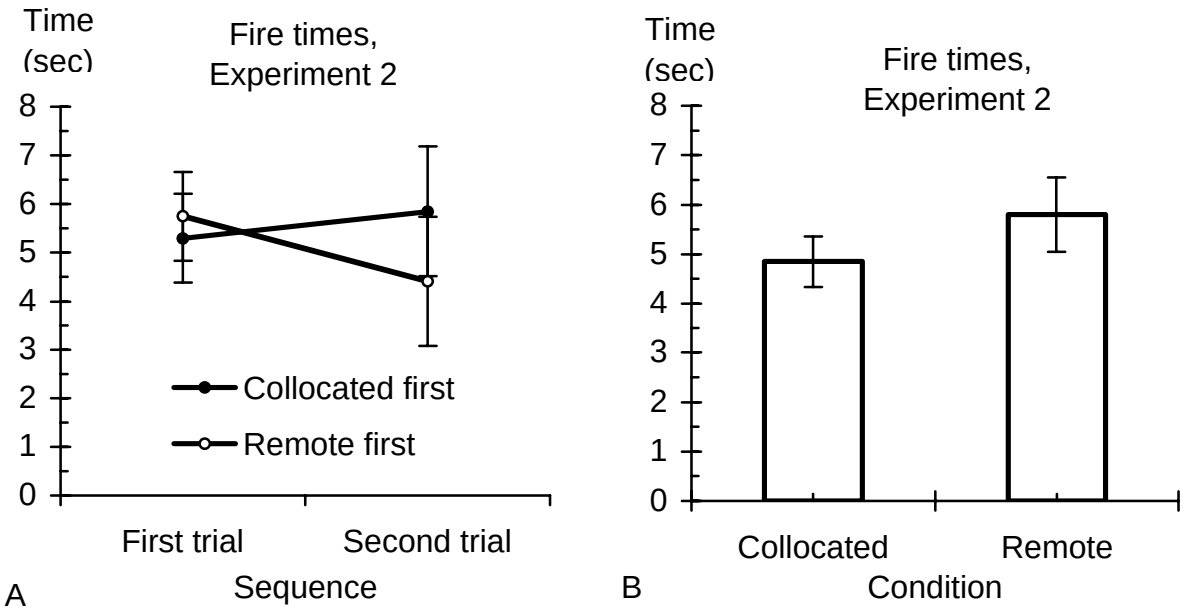


Figure 5 Average response times to commands to fire (A) and the distribution of differences in response times between the remote and collocated conditions (B), Experiment 2.

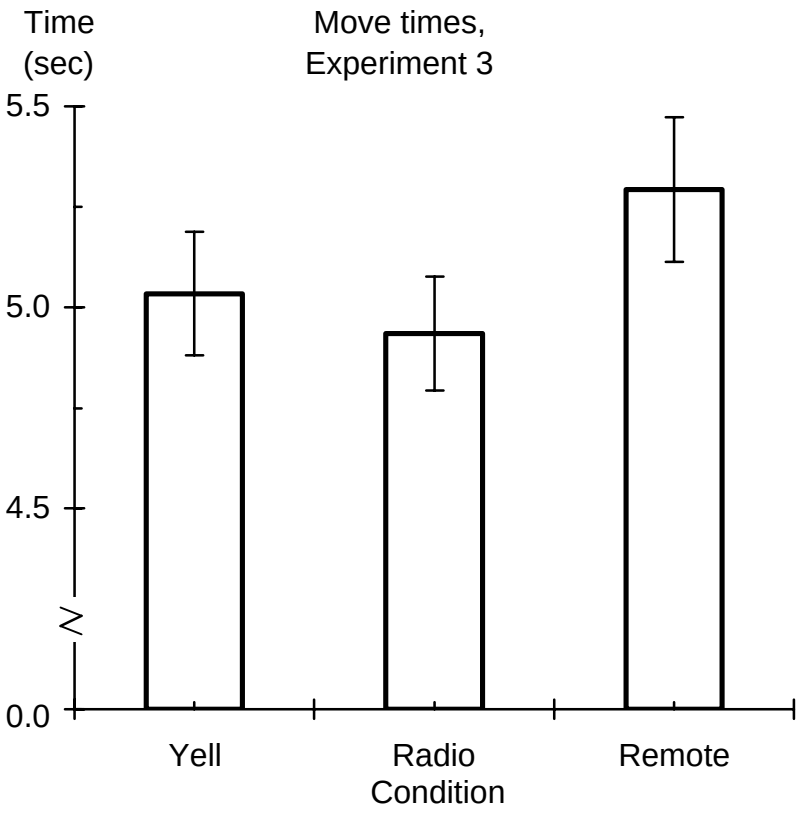


Figure 6 Average response times to commands to move, Experiment 3.

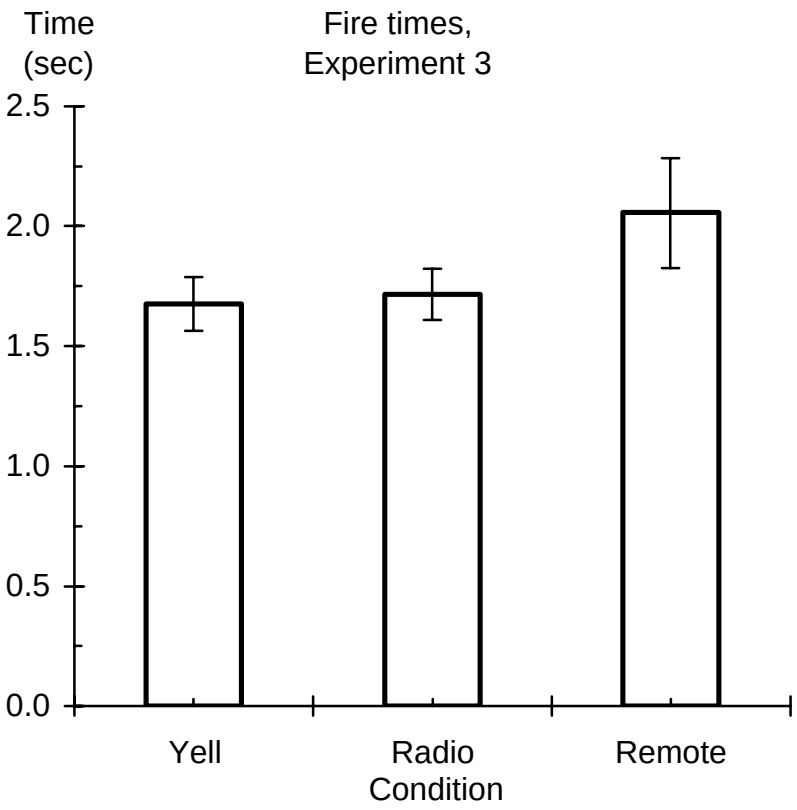


Figure 7 Average response times to commands to fire, Experiment 3.

Running head: TIME TO CONTACT IN REPRESENTATIONAL SPACE

Time-To-Contact in Representational Space

Kip Smith

Linköping Institute of Technology

P. A. Hancock

University of Central Florida

Address correspondence to Kip Smith, MURI-OPUS Research Laboratory, Industriell arbetsvetenskap, Institutionen för Konstrucktion och produktionsteknik, Linköpings universitetet, 581 83 Linköping, Sweden; e-mail: kipsm@ikp.liu.se. Phone: (011) 46 13 282764 Fax: (011) 46 13 282579

Note to the editor and reviewers: The international aviation community generally does not use the metric system. Altitudes are almost always reported in feet and velocities in knots (nautical miles per hour). The units used in this paper are those in use around the world.

Abstract

Accepted theory maintains that humans detect and avoid collisions through the direct perception of time-to-contact. Here, we report the first results that demonstrate the perception of time-to-contact in a purely representational rather than a directly perceived space. In simulations of en-route air traffic, experienced airline pilots detected and avoided collisions using information provided solely by a flat-panel, god's-eye view display of traffic information. Their behavior reveals an invariant relationship between the time of conflict detection and the separations and relative velocities of aircraft in the display. The invariant serves as an avoidance heuristic for the airborne self-separation of aircraft. This insight into human behavior should have an immediate and practical application in the next generation of the National Airspace System.

Time-To-Contact in Representational Space

To avoid the adverse consequences of collision and to facilitate affective touch, mobile agents employ their various sensory capacities to detect time-to-contact directly (Gibson, 1954, 1958; Lee, 1974). For example, human and monkey infants use a visual capacity to detect and attempt to avoid looming objects when only a few months old (Kayed & van der Meer, 2000; Yonas, Pettersen, & Lockman, 1979; Schiff, Caviness, & Gibson, 1962). The information that guides such behavior at all ages is the rate of expansion of the image of a moving object on the retina or of an object on a moving retina (Stoffregen & Riccio, 1990). Gibson, Lee, and others have argued that the rate of retinal expansion uniquely and immediately affords information of time-to-contact which obviates the need for any symbolic computation to avoid collisions (Bootsma, 1988; Bootsma & Oudejans, 1993; DeLucia, 2003; Gibson, 1979/1986). By this *direct* or *ecological* account, visual time-to-contact requires the immediate presence of an object within the unaided range of visual perception and the direct detection of the flow of light reflecting off that object (Lee, 1976; McLeod & Ross, 1983). Similar principles could apply to other sensory capacities as well.

A formidable challenge to this account is presented by the evident success of human operators like air-traffic controllers who routinely rely on two-dimensional displays that transform time-to-contact information into symbolic form. Clearly there is more to time-to-contact than just the stimulus dynamics alone (and see Hancock & Manser, 1997). In this work, we report an empirical demonstration of a time-to-collision percept in a representational space vastly different from that posited by accepted ecological theory.

An alternative to the ecological account for apprehending or comprehending time-to-contact or both is the computational method. In this account, estimates of distance and velocity

are used as bases for subsequent mental calculation to provide time-to-contact estimates. While the direct method is largely devoid of any cognitive or central mediation, the computational method is often called *indirect* because of its need for an additional, goal-oriented transformational step.

As a central pillar of the ecological perspective on behavior, the direct availability of the information is a powerful argument that the sensory systems are *tuned* to the ambient stimulus array. Theorizing about the fit between sensory systems and the environment in which they evolved suggests that complex cognitive transformations of sensory data are unnecessary elaborations upon what are much simpler and parsimonious explanations. The rigid ecologist would like to view the direct perception of time-to-contact as necessary, sufficient, and exclusive.

Such strong protestation may be premature. Simply because the adaptive nervous system can extract time-to-contact directly from ambient displays does not imply the level of exclusivity hoped for by an adamant theoretician. It is clean that modern day technology provides a spectrum of representational displays (e. g., the speedometer, the altimeter) that integrate and render sensory data in ways that diverge radically from that afforded by the natural environment. Although some design theorists advocate for the inverse, that is, for basing design principles for displays on natural constraints (e. g., Vicente, 2004), such a recommendation has yet to be widely implemented. As a result, many displays provide an ad hoc collection of representations of systems at various levels and dimensions. Consequently, it is crucial to understand if perceptual primitives such as time-to-contact work in domains filled with representational artifacts. Commercial aviation is an exemplar of such a domain.

Since the advent of the air traffic control system in the mid-1950's, controllers have been using composite radar images to detect and avert impending midair collisions. The recent midair incident over Ueberlingen, Germany provided tragic evidence of the consequences of failure to detect impending collision (Flottau, 2002)[1]. The repeated breakdowns of the information systems that support centralized air traffic control systems raise the specter of similar catastrophes in the United States' National Airspace System (NAS). With this in mind, the Federal Aviation Administration (FAA) is exploring a distributed control system known as *free flight* (RTCA, 1995). In free flight, commercial pilots in en-route airspace (above 19,000 ft.) communally act as their own airborne traffic control system. The technology to support such a distributed control system is envisioned to use cockpit displays of traffic information (CDTI) relying on global positioning systems. A CDTI is a computer-driven flat-panel display that uses symbols to represent the locations and velocities of aircraft. What remains unresolved is whether and how such symbolic representations of the airspace support the pilot's genetic inheritance of time-to-contact and, in this domain, afford the timely and effective detection of incipient collisions.

---

[1] In fairness to the controller, this collision was precipitated by a conjunction of events and a clash of cultures in addition to his tardy detection of the impending collision. It is generally inappropriate to brush off such incidents merely as the product of "human error" (see Dekker, 2002).

Method

Participants

Ten highly-experienced and currently-certified commercial airline pilots (flight hours: mean 9,177, range 6,000 to 24,000; age: mean 51.9, range 38 to 63) volunteered to participate[2]. After receiving instructions, signing an informed consent form, and becoming familiar with the simulator, its cockpit, and its displays in a series of practice scenarios, the pilots individually flew a sequence of twelve realistic but challenging en-route flight scenarios.

Apparatus

The part-task glass-cockpit simulator replicates the instrumentation and dynamics of a Boeing 757 aircraft (Klinge, Smith, & Hancock, 1997). Instrumentation included, from left to right on the cockpit panel, a primary flight display, the CDTI, and a flight management system with a multi-control display-unit keypad. With the exception of the custom-designed CDTI shown in Figure 1, the displays and keypad had the look and feel of standard equipment in a commercial Boeing 757. The computer-generated CDTI presented a top-down view of the airspace surrounding the pilot's aircraft and was the pilot's only source of traffic information. The CDTI was displayed on a 15 cm x 22.5 cm screen at the center of the cockpit panel. A typical image in forward-view mode is shown in Figure 1.

The apex of the large triangle at the center of three semicircular range rings (or in the center of the display in full-view mode) represented the location of the pilot's aircraft. The

---

[2] Participants were certified to fly one or more of the following glass-cockpit aircraft types: Airbus 320, DC9/10, MD80/90, Boeing 474-400, 757, and 767.

pilot's heading was always shown in the noon position. The bearing of the heading was shown both numerically at the noon position and with 60° of arc of a compass rose. Symbols representing ground stations (not shown in Figure 1) flowed down the screen at the apparent velocity (*ground speed*) of the pilot's aircraft. The next waypoint[3] on the pre-programmed (or pilot altered) flight plan (e.g., LIT – Little Rock, Arkansas) and the estimated time to that waypoint (28 minutes) were shown at the top of the display. The numbers in the bottom right corner of the display indicated the aircraft's ground speed in knots and altitude in feet.

The positions of other aircraft (traffic) were represented by triangles with short intent vectors indicating their direction of travel. Associated with each aircraft was a *datatag* shown in standard FAA format. For example, consider the aircraft designated NWA 656 in the upper left of Figure 1. The top line of text in the datatag indicates the flight's callsign (NWA 656); the bottom line of text indicates its ground speed in knots (nautical miles per hour), heading in degrees divided by 10 (e.g., 32 indicating 320°), and flight level (310, altitude in 100's of feet). While cryptic to an outsider, this format and mix of non-metric units were highly familiar to the participants. Traffic flowed across the screen at their relative velocities and headings with respect to the pilot's aircraft.

In the scenario depicted in Figure 1, the pilot's aircraft is approximately 10 nautical miles behind an aircraft with the callsign USA 311. Both are flying at 296 knots at an altitude of 31000 feet. NWA 656, also at 31000 feet, will cross the pilot's path more than 20 but less than

---

[3] A waypoint is a ground station with an FAA-maintained radio beacon used to establish and maintain a flight plan.

30 miles ahead. Since all aircraft are traveling at approximately 8 miles per minute, the pilot's aircraft and NWA 656 are likely to come into conflict in approximately 3 minutes.

The CDTI supported a flight tracking mode which the pilot could turn on or off. When on, an bright line extended from each aircraft in the display to trace the route of its flight plan. This mode added considerable clutter to the display but was used by the pilots from time to time to ascertain the intent of specific aircraft.

Simulations ran via a network of Pentium-based computer workstations running Linux OS. Software enabled the pilot to navigate the simulated airspace using instrument flight rules and procedures and to maintain or deviate from a preprogrammed flight plan using the flight management system and keypad.

The size of the triangles representing other aircraft in the airspace remained constant throughout the session and zoom settings of the display. Thus, the optic array with information about time-to-contact contained none of the cues central to accepted theory (e.g., Gibson, 1979/1986; Lee, 1974, 1976).

Procedure and Design

Pilots were instructed to adhere to the FAA's rule for separation between their aircraft and all other aircraft. This rule has two parts, a minimum horizontal distance between aircraft of 5 nautical miles and a minimum vertical distance of 1,000 feet. This thin disk defines a *protected zone* around each aircraft. In the current system of ground-based air traffic control, a controller receives a stiff penalty of demerits and retraining if two aircraft in his or her sector of the airspace violate each others' protected zones.

The pilots were free to make all decisions about routing and separation without the support or intervention of air traffic control. In 11 of the 12 scenarios, pilots encountered realistic but challenging traffic conditions that were designed to generate a conflict at a scheduled time if the pilot took no action to avoid it. The twelfth scenario, a control condition, contained no scheduled conflict.

Table 1 shows the incomplete block design of the 12 en-route scenarios. The design manipulated two dimensions of airspace complexity – traffic count at three levels and relative bearing of traffic at five levels. Traffic count, a common metric of airspace complexity (e.g., Hopkin, 1980; Smith, Scallen, Knecht, & Hancock, 1998) was defined the total number of aircraft that came within 150 km of the pilot's aircraft prior to and during the scheduled conflict. Relative bearing is defined with reference to the direction of flight of the pilot's aircraft. Thus, traffic with a relative bearing of 0° is flying in the same direction as the pilot. Faster aircraft directly behind the pilot flying at a relative bearing of 0° pose the threat of an *overtaking* conflict. Traffic with a relative bearing of 180° approaches the pilot head-on. Roughly half of the traffic with intermediate values of relative bearing crossed the pilots path from the left and half from the right.

In a series of shake-down flights, the pilot who volunteered to serve as our subject-matter expert commented that the 0° values of relative bearing (being overtaken) and 180° value (head-on collision at the same altitude) were unrealistic but challenging. To reduce the time participants would have to spend in the simulator, we retained only one instance of each of these bearings. The 11 conflict scenarios developed violations of the FAA's minimum separation rule within six minutes if the pilot followed the preprogrammed flight plan.

Dependent measures

Both performance and process data were collected as the pilots flew the air traffic scenarios. Performance measures included the locations and velocities of all aircraft, all keystrokes and button pushes. Our simulator records, time stamps, and writes to a binary output file the basic data needed to replicate the experimental session. These data include (a) the initial vector, e.g., starting point (x, y, z) and velocity, of all aircraft, (b) the time and nature of any change to that vector, and (c) the time of every button push or dial setting. Among the recorded data are the number of maneuvers executed and the number of conflict situations encountered. From these data, a post-processing algorithm calculated the separation and relative velocity for each aircraft with respect to the pilot's aircraft at approximately 5 Hz throughout each scenario.

The process data were the pilots' concurrent verbal reports. Pilots were asked to think aloud as they navigated through the traffic. These reports were captured on both audio- and videotape and transcribed using the 'pause-and-clause' technique: every segment of speech became a line in the transcript. Each clause was synchronized with the videotape and annotated with a time-stamp.

We then analyzed the transcripts to determine when a pilot first indicated he had detected an impending conflict. The criterion for detection was a clause containing some reference to another aircraft, usually the pronoun "he" or the callsign of an aircraft. Many detection statements were prefaced by an explicative and contained a self-referential statement. The time stamp corresponding to a detection statement we call the *detection time*. The target aircraft that motivated the pilot's detection of an impending conflict and the subsequent sequence of actions was determined by computer playback of the session. The post-processing algorithm provided the separation and relative velocity for each target aircraft at its detection time. The pilots'

detection times and the corresponding separations and relative velocities of the pilots' aircraft and the conflict aircraft are the data used in the regression analysis.

Results

Grouping across participants and scenarios, the pilots were able to maintain the FAA separation criteria in 84 of the 110 conflict scenarios. In 26 scenarios, the pilots did not detect the impending collision before they violated the FAA's rule for minimum separation between aircraft. The relatively high percentage of successful resolutions (84/110 = 76%) reveals that the pilots were generally able to detect and resolve conflicts using information from the CDTI shown in Figure 1. The presence of failures to detect conflicts says more about the difficulty of our scenarios than it does of the ability of the pilots. Our focus here is on the determinants of their success at extracting time-to-contract information from the flat-panel CDTI.

Process data

In 28 of the 84 scenarios in which the pilots detected and maneuvered to avoid an impending conflict, the pilot did not verbalize with sufficient clarity for us to identify either the detection time or target aircraft unambiguously. The analysis is based on the 56 remaining successful detections.

The following is a excerpt from one of the pilot's verbal reports containing an example of a successful detection of time-to-contact with another aircraft. The pilot is approaching the LANDO waypoint.

4:18 I'm looking at American 333 (turns on flight tracking mode)
4:23 He's got LANDO as well (turns off flight tracking mode)
4:28 I don't have ETA for him there but he's doing 410 in ground speed and he's got about 60 miles to go so
4:35 So, yeah, probably about the same time I'm going to get there
4:40 I'm climbing

In this passage the detection time is 4 minutes 35 seconds into the session when he says "Yeah." In the preceding 17 seconds the pilot has identified the target on the CDTI, inferred that the target is headed for LANDO as well, inspected its datatag for its ground speed (and presumably altitude), and used the range rings on the display to infer the target's distance from LANDO. There is little evidence for explicit calculation. Rather there is a holistic assessment that his aircraft and the target aircraft are likely to be in the same place (and at the same altitude) at "about the same time," not a good thing. The pilot then resolves the conflict by changing altitude.

In the second example, from a different pilot, there is minimal evidence for mental calculation. The only explicitly numerical consideration is altitude. The pilot's correct inference that "Continental 57 is a problem" comes from inspection of the target's data tag and flight track.

0:17 What's going on here now? I'm point 8 mach 33000 feet heading 020
0:26 Looks good. …
2:28 There's a Delta at 031
2:36 Yep, and he's – let's open the scale a bit (changes the radius of the maximum range ring to increase the field of view)
2:39 so I can get a little farther away
2:41 Continental, let's say he's 330
2:46 He's at 31000
2:48 He's at 33000 so I want to track him
2:55 I want to track Continental so let me look at the status of the traffic
3:01 Continental 57 and I want his path on (turns on flight tracking mode)
3:07 Ok, I'm going to keep
3:14 He's at 31 I'm at 33 the Delta
3:17 Continental 57 is a problem
3:28 Well, this is like real flying (turns off flight tracking mode)
3:36 My poor passengers have been having a terrible time (laughs)
3:41 Ok, Delta's going behind me, Continental is in front of me
3:51 He's at the same altitude
3:53 Ah, lets see what scale I'm on – 30
3:58 I don't like that
4:00 I think I'm going to descend
4:05 Lets see, he's at 31, I'm at 33, I don't want to descend, I'm going to climb
4:12 33, I don't like that either

4:17 Ah, I think I will turn behind him

More than a minute passes between the detection time (3:17) and the final selection of a maneuver to resolve the conflict. This suggests that the inferred time-to-contact provides the pilot sufficient leeway to investigate a series of options. The pilot's concern rests not with detecting time-to-contact but with what to do about the inferred conflict. The pilot extracts information about time-to-contact directly from the display. His decision making focuses on flying the aircraft.

Performance data

A graph of the performance data reveals the nature of the information in the CDTI that the pilots use to make statements like "So, yeah, probably about the same time I'm going to get there" and "Continental 57 is a problem." Figure 2 is a phase plane representation of en-route air-traffic dynamics (Phatak & Bekey, 1969; Smith & Hancock, 1995). A phase plane is a state space in which one of the axes is the time derivative of the other (Jagacinski & Flach, 2003). In Figure 2 the vertical axis is separation, the distance between the locations of the pilot's aircraft and the target aircraft in nautical miles. The horizontal axis is its derivative, velocity in knots. Here, velocity is the relative velocity of the two aircraft, the sum of the components of their velocities in the direction of their separation.

Each circle in Figure 2 represents the separation and relative velocity between the pilot's aircraft and a target aircraft at its detection time for one of the 56 conflicts that were clearly detected and successfully resolved. Many of the data overlap. Circles near the vertical axis represent conflicts in an overtaking scenario that unfolded slowly. Circles at the far right represent conflicts with aircraft approaching nearly head-on and, accordingly, with high relative

velocity. Circles close to the horizontal axis represent conflicts that occurred with in two minutes of the beginning of the scenario.

The data shown in Figure 2 are collapsed across the manipulations of traffic count and relative bearing. In spite of this large diversity of conflict opportunity and the variation which results from intrinsic individual differences, the data exhibit a significant linear trend. Least-squares linear regression defined a line with an intercept of 5.8 nautical miles and a slope of 0.046 hours = 2.75 minutes ($r^2 = .52$, $n = 56$). The 95% confidence interval for the line, shown with dashed curves in Figure 2, indicates that the Y-intercept is statistically indistinguishable from the 5-mile FAA criterion for minimum horizontal separation.

Discussion

The slope and intercept of the regression line in the state space of Figure 2 indicate that commercial pilots were able to use the CDTI shown in Figure 1 to monitor traffic and to detect conflicts roughly three minutes prior to the time when they would violate the FAA's criterion for minimum horizontal separation. Given the flight characteristics of the Boeing 757 (and all other current commercial airliners), the three minute cushion provided sufficient lead-time to take corrective action and therefore to maintain the 5-mile separation criterion.

We interpret this regression line as parsing the state space into two regions, a region associated with monitoring and a region associated with maneuvering. Above the threshold, pilots actively monitor the CDTI and continue with their current flight plan. Below the threshold, pilots explore, select, and execute maneuvers designed to resolve an impending conflict. Thus, the threshold represents an invariant in the pilot-airspace system that triggers task-relevant action across participants and situations. The invariant has units of time, the time-to-contact with another aircraft’s protected zone.

The invariant in the state space for free flight defines two classes of thinking in action, a hallmark of adapted individuals (Clarke & Crossland, 1985; Suchman, 1987). Like adapted individuals everywhere, the pilots' thinking in action is driven by the elemental information in their environment. Here, that information is separation and relative velocity as afforded by a flat panel display. This result is the first demonstration of which we are aware of time-to-contact afforded by a representational space.

The constant in the regression equation is statistically equivalent to the 5-mile FAA criterion for minimum separation. This constant was initially taken as evidence that pilots were demonstrating a successful adaptation to the central organizational constraint on navigation in en-route airspace. By this account, the regression line has a 5-mile intercept because the pilots were accustomed to heeding the FAA's organizational constraint. Presumably, pilots were acting in a manner consistent with rules and lessons learned during extensive training in simulators and thousands of flying hours.

Scrutiny of the tasks performed by commercial pilots suggests an alternative account. In the current structure of the National Airspace System, air traffic controllers, not commercial pilots, are given the task of ensuring that aircraft heed the FAA's arbitrary 5-mile separation criterion. Prior to our study, the pilots whose behavior is described in Figure 2 had never been given the task of ensuring separation. This inverts the analysis. We suggest that the adoption of a 5-mile limit (rather than a 3 mile or 10 mile limit) was a tacit acknowledgment of the information that pilots heed in order to take account of others in the airspace. The arbitrary organizational constraint may itself be a manifestation of the invariant that directs behavior in the environment it regulates. The invariant we have demonstrated apparently pervades the distributed consciousness of the National Airspace System.

Systems and organizations can be divided into two categories: those that demand error-free performance and those that stress efficiency (Weick & Roberts, 1993). The National Airspace System is an exemplar of the former category. All of the partners in commercial and civil aviation demand continuous operational reliability because the consequences of error are catastrophic. The transition to free flight in en-route airspace will not change this demand. Free flight will be realized if and only if it is implemented in a manner that fits the adaptation of its participants. To this end, regulators must heed how pilots interact with the airspace. That experienced pilots detect and act to resolve impending collisions with a base percept of an envelop with a radius of 5 nautical miles and a 3 minute cushion strongly supports the contention that this is a learned behavior erected on some form of an innate capacity. This behavior can be used to facilitate collision avoidance in commercial and civil aviation and to form the basis for regulations governing the FAA's vision of free flight. The threshold in the separation-velocity state space is the natural element of the model sought by the cognitive engineers of free flight.

The implications for the competing theories of time-to-contact are equally intriguing. Pilots report no conscious effort to perform calculations derived from distance and velocity. Rather, the decision to alter a flight path appears to emerge from an implicit recognition of potential conflict. We suggest that every individual erects a *zone of safe passage* around themselves and similarly around themselves and any vehicle they control. This construct, first promulgated by Gibson and Crooks (1938) for automobile control, implies that the time-to-contact tolerance of each respective zone is contingent upon the dynamics of the vehicle involved, whether that vehicle is the individual's own body, an automobile, or a commercial jetliner. The results presented here reveal an implicit adherence to a collective recognition that the three minute interval is a crucial threshold in commercial aviation operations. We would

expect to see similar constants in driving and unaided locomotion, although of course the value of time cushion would certainly be reduced. That individuals manifest this protective envelope, predicated only upon an arbitrary representational display, is both a unique and encouraging finding.

Acknowledgments

This work was funded by the Federal Aviation Administration through Grant 93-G-048. P. A. Hancock and K. Smith were the Principal Investigators. Dr. Tom McCloy was the technical monitor. We thank Mr. Jeffrey Lewin for assistance with data reduction, Mr. Guy Smith for his guidance, Northwest Airlines for access to pilots, and Honeywell International for the Boeing 757 multi-control display-unit keypad. The views expressed are those of the authors and do not necessarily represent those of the FAA or its agents.

References

Bootsma, R. J. & Oudejans, R. R. D. (1993). Visual information about time to collision between two objects. Journal of Experimental Psychology: Human Perception and Performance **19**, 1041-1052.

Bootsma, R. J. (1988). The timing of rapid interceptive actions: Perception-action coupling in the control and acquisition of skill. Amsterdam: Free University Press.

Clarke, D. D. & Crossland, J. (1985). Action systems: An introduction to the analysis of complex behavior. New York, NY: Methuen.

Dekker, S. W. A. (2002). The field guide to human error investigations. Aldershot, UK: Ashgate.

DeLucia, P. (2003). Information integration in judgements of time to contact. Quarterly Journal of Experimental Psychology A **56A**(7), 1165-1189.

Flottau, J. (2002). Human factors role cited in German crash. Aviation Week and Space Technology (Sept. 9), 44.

Gibson, J. J. (1954). The visual perception of visual motion and subjective movement. Psychological Review **61**, 304-314.

Gibson, J. J. (1958). Visually controlled locomotion and visual orientation in animals. The British Journal of Psychology **49**, 182-194.

Gibson, J. J. (1979/1986). The ecological approach to visual perception. Boston: Houghton-Mifflin.

Gibson, J. J., & Crooks, L. (1938). A theoretical field analysis of automobile driving. American Journal of Psychology **51**, 453-471.

Hancock, P. A. & Manser, M. (1997). Time-to-contact: More than tau alone. Ecological Psychology **9**, 265-297.

Hopkin, V. D. (1980). The measurement of the air traffic controller. Human Factors **22**(5), 547-560

Jagacinski, R. J. & Flach, J. M. (2003). Control theory for humans: Quantitative approaches to modeling performance. Mahwah, NJ: Lawrence Erlbaum Assoc.

Kayed, N. S., & van der Meer, A. (2000). Timing strategies used in defensive blinking to optical collisions in 5- to 7-month-old infants. Infant Behavior & Development **23**, 253-270.

Klinge, J., Smith, K., & Hancock, P. A. (1997). DATIDS - The University of Minnesota distributed air-traffic information display simulator. Proceedings of the 9th International Symposium on Aviation Psychology, Columbus, Ohio.

Lee, D. N. (1974). Visual information during locomotion: In R. B. McLeod & H. Pick (Eds.), Perception: Essays in honor of J. J. Gibson (pp. 250-267). Ithaca, NY: Cornell University Press.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. Perception **5**, 437-459.

McLeod, R. B., & Ross, H. E. (1983). Optic flow and cognitive factors in time-to-contact estimates. Perception **12**, 417-423.

Phatak, A. V., & Bekey, G. A. (1969). Decision processes in the adaptive behavior of human controllers. IEEE Transactions of Systems, Science and Cybernetics, SSC-5, 339-352.

RTCA (1995). Free Flight Implementation, Final Report of Task Force 3. Washington, D.C.: RTCA.

Schiff, W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of "looming." Science **136**, 882-983.

Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally-directed consciousness. Human Factors **27**(1), 137-148.

Smith, K., Scallen, S. F., Knecht, W. R., & Hancock, P. A. (1998). An index of dynamic density. Human Factors **40** (1), 69-78.

Stoffregen, T. A. & Riccio. G. E. (1990). Responses to optical looming in the retinal center and periphery. Ecological Psychology **2**(3), 251-274.

Suchman, L. A. (1987). Plans and situated actions: The problem of human machine communication. New York, NY: Cambridge University Press.

Vicente, K. (2004). The human factor: Revolutionizing the way people live with technology. Scarborough, ONT: Routledge

Weick, K. E., & Roberts, K. H. (1993). Collective mind in organizations: Heedful interrelating on flight decks. Administrative Science Quarterly **38**, 357-381.

Yonas, A., Pettersen, L., & Lockman, J. J. (1979). Young infants' sensitivity to optical information for collision. Canadian Journal of Psychology **33**, 268-276.

Table 1

The manipulations of traffic count and relative bearing in the 12 en-route air traffic scenarios

Traffic count	Relative bearing of conflicting traffic					No conflict
	0°	45°	90°	135°	180°	
6		X	X	X		
9	X	X	X	X	X	X
16		X	X	X		

Figure captions

Fig. 1. The cockpit display of traffic information (CDTI) used in the experiment. For clarity of presentation, the colors have been inverted and the display set to forward-view mode at the smallest maximum range setting, 30 nautical miles. In this mode the location of the pilot's aircraft is the apex of the large triangle in the middle of the bottom of the display. The pilot could change the display to a 360° view, putting his aircraft at the center of concentric range rings. The pilot could also increase the maximum range incrementally to as much as 300 nautical miles. The size of traffic symbols and their datatags remained constant across ranges. At large ranges, traffic and their datatags often overlapped. The symbology is discussed in the text.

Fig. 2. The state space for free flight showing the information that afforded successful resolution of impended collisions and the linear least-squares regression fit to those data. The vertical axis is the separation between the pilot's aircraft and another aircraft seen in the CDTI of Figure 1 at the time when the pilot detected an impending conflict situation. The horizontal axis is the rate of change of separation between the aircraft, that is, their relative velocity. Curves represent ±95% confidence interval for the regression line. Plotted points may overlap. The slope of any line in the state space has units of time. The Y-intercept of a line indicates a critical distance between aircraft and its slope indicates the time when that separation will occur.

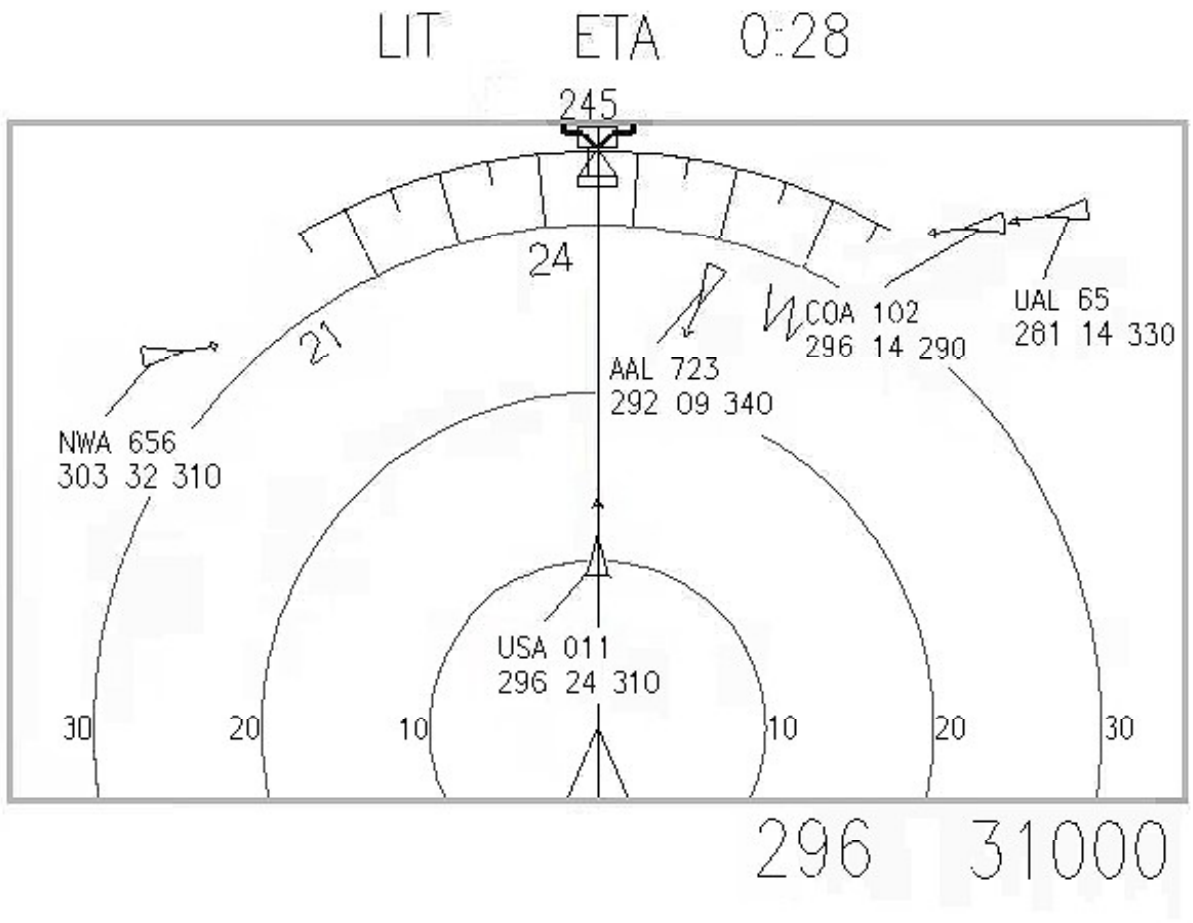

LIT ETA 0:28
245
24
21
NWA 656
303 32 310
AAL 723
292 09 340
COA 102
296 14 290
UAL 65
281 14 330
USA 011
296 24 310
30
20
10
10
20
30
296 31000

Figure 1

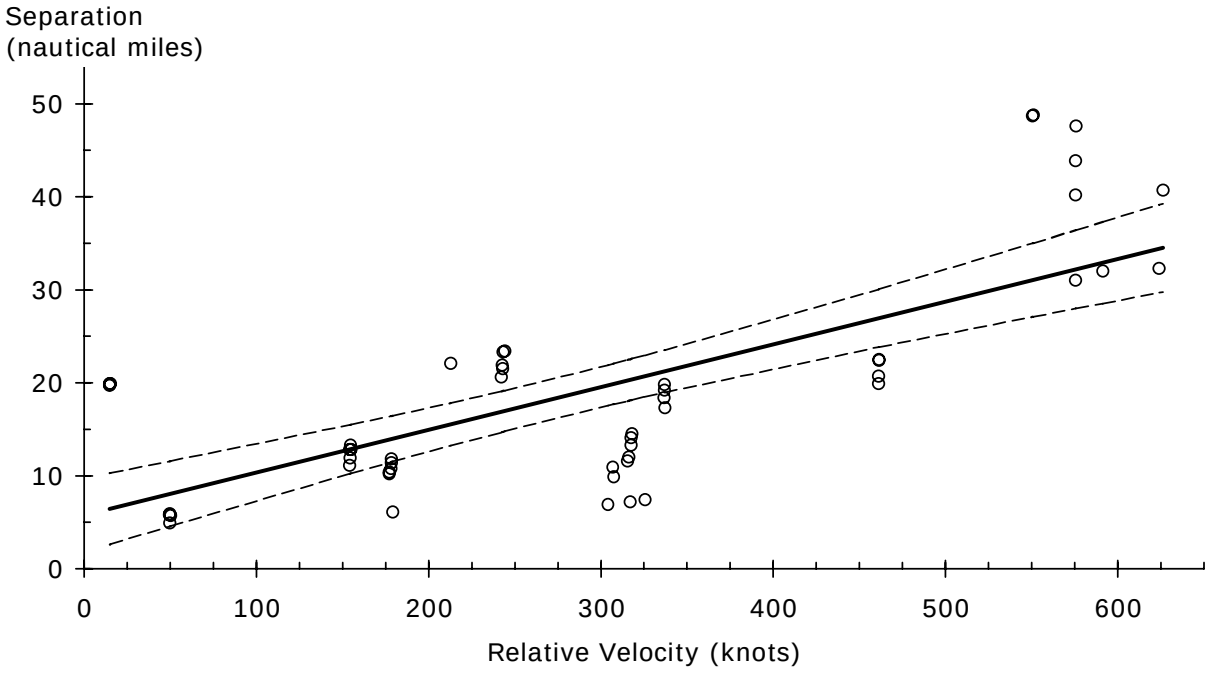

Separation
(nautical miles)
50
40
30
20
10
0
0
100
200
300
400
500
600
Relative Velocity (knots)

Figure 2

Running head: HUD USE AND LANDING PERFORMANCE

HUD use improves landing performance in restricted visibility

Örjan Goteman

Scandinavian Airlines

Kip Smith and Sidney Dekker

Linköping Institute of Technology

Abstract

The operational community has assumed that using a head-up display (HUD) instead of conventional head-down displays will increase accuracy and safety during approach and landing. This paper reports a pair of experiments that test that hypothesis. Using a Boeing 737-700 full-motion simulator flown by commercial airline pilots, we explored the effects of (1) HUD use, (2) ambient visibility, and (3) length of approach lighting on the size and location of the touchdown footprint. HUD use reduced the width of the touchdown footprint in all tested visibility and lighting conditions including visibility below the minimum allowed. HUD use had no effect on the length of the touchdown footprint. Based on these empirical data, the minimum visibility for approaches using HUD could be set lower than for approaches without a HUD.

HUD use improves landing performance in restricted visibility

This paper reviews two experiments designed to assess the impact of commercial airline pilots' use of a head-up display (HUD) on approach and landing performance under reduced runway visual range (RVR) conditions. We begin by reviewing previous research on HUD use and by describing the effect of how ambient RVR affects approach and landing operations. The review motivates a pair of hypotheses regarding the impact of HUD use on the size of the landing footprint. The experiments and their findings are discussed in turn with particular emphasis on the mechanisms behind the effects and the practical implications for approach and landing in restricted visibility.

HUD use in approach and landing operations

The HUD projects computer-generated aircraft flight-path and energy symbols onto a transparent screen. Pilots can concurrently view the HUD symbology and the surrounding world. Commercially available HUDs usually replicate the information made available by the pilot's conventional flight instruments, showing aircraft attitude, speed, altitude and heading. In addition, HUDs typically contain a flight path symbol. Commercial aircraft equipped with a HUD usually have it installed in front of the left pilot (captain) only. Conformal HUD displays preserve a 1:1 relationship between the visual angles of displayed symbols and their real world referents. A conformal HUD with a flight path symbol can thus instantaneously show the pilot where the aircraft is going relative to the surrounding world. In contrast, a pilot relying on conventional flight instruments must infer the aircraft's flight path based solely on his/her perception of optical flow (Gibson, 1986).

Comparisons between HUD and head-down displays in manual flight have found that conformal HUD use improved track, speed, and altitude maintenance (Lauber et al., 1982; Martin-Emerson & Wickens, 1997). The civil operational community assumed that these HUD performance advantages over conventional head-down instrumentation could reduce the number of approach and landing incidents and accidents (FSF, 1993). Specifically, the commercial operational community assumed that HUD use could mitigate two well-documented problems associated with approach and landing: (1) visual approaches to runways without radio navigation aids or with unreliable navigation aids, and (2) the transition from instrument to external visual cues for landing in low visibility (e.g., Newman, 1995). Apart from military research reported by Hoerner, 1979 (as cited in Newman, 1995) reporting a threefold reduction in landing accidents on aircraft carriers when pilots used a conformal HUD with a flight-path symbol, these assumptions appear to have gone untested in the commercial aviation.

The presumed sources of the advantage in flight-tracking performance for the HUD are that it eliminates the need for the pilot to move his or her gaze from head-down instruments to the outside world to look for maneuvering cues (Stuart, McAnally, & Meehan, 2003) and minimizes scanning requirements (Martin-Emerson and Wickens, 1997). One consequence of the transition from heads-down to the outside world is the need for visual accommodation (e.g., the visual depth of field changes less than a meter to infinity or vice versa). Because conformal HUD symbology is focused at infinity, HUD use eliminates the need for and time demand of visual accommodation, simplifying the pilot's task.

Empirical support for this assumption and for HUD use is relatively weak. Iavecchia and Roscoe (1988) reported observing over-accommodation (when the pilot's focus turns to the HUD symbology) despite the collimated optics when pilots used a HUD. This lead them to predict

poorer touchdown control. Specifically, they predicted that over-accommodation would contribute to longer and shallower landings with HUD. However, it seems that they did not test this hypothesis. Wolffsohn, Edgar, Stone, Williams, and McBrien (1999) defend HUD by reporting over-accommodation of small magnitudes in subjects using HUD in a cognitive demanding task. This led them to suggest that the observation of over-accommodation is a product of the relatively high cognitive load of the task itself and is not due to the HUD per se.

The same studies that established a performance advantage for HUD have also shown signs of a potential weakness: attentional capture, also known as cognitive tunneling or cognitive fixation when using the HUD. Several studies have reported a slower response to unexpected outside-world events such as runway obstacles when pilots were using a HUD (Weintraub, Haines, & Randle, 1985; Wickens and Long, 1995) or a performance trade-off from flight path to altitude maintenance when using HUD symbology (Sanford, Foyle, McCann, & Jordan, 1993). These findings are consistent with attentional capture by the HUD symbology at the cost of to unexpected information in the optic flow from the environment (Martin-Emerson & Wickens, 1997). Attentional capture describes maladaptive behavior in which the focus of attention becomes highly restricted and searches for information become limited and largely confirmatory. Such behavior was found by Fadden, Ververs and Wickens (1998) who showed that pilots using HUD were quicker to detect expected events and slower to detect unexpected events than pilots relying on head-down displays. However, Martin-Emerson and Wickens (1997) found no evidence for attentional capture when that HUD symbology was truly conformal to the optic flow from outside world. For a review of attentional capture and HUD use see Stuart, McAnally and Meehan (2003).

The clear advantage at flight-path tracking afforded by a conformal HUD has overshadowed the possibility that HUD use promotes attentional capture and has encouraged the operational community to install HUDs in commercial flight decks. Commercial pilots have been using HUDs for approximately 20 years both en-route and during approaches and landings. However, it appears that there have been no studies of the purported advantages of HUD use during approach operations.

Approach operations in reduced visibility

The majority of approaches (the segment of flight immediately before touchdown) performed in US and European civil air transport operation are conducted as Category I instrument landing system (ILS) approaches. For category I approaches two critical factors when making the decision of whether or not to land are the decision height and the runway visual range (RVR). Decision height refers to the aircraft's vertical distance above the runway threshold where the pilot must make the decision to land or make a go-around. The RVR is a measure of horizontal visibility defined by the number of visible approach and runway lights in the ambient atmospheric conditions. If the RVR is too low at the decision height, the pilot will not be able to see some or any of the approach lights and must make a go-around. Although the resulting missed approach is part of normal operations (and formal procedures), it adds an undesired additional risk (ICAO, 2003).

To commence a Category I ILS approach, the RVR measured at the runway must not be less than 550 m (International Civil Aviation Organization [ICAO], 2001). An RVR of 550 m will allow the pilot to see a visual segment of the ground that contains enough of the runway approach lights to judge the aircraft's lateral position, cross track velocity, and position in roll when the aircraft is at decision height (Joint Aviation Authorities [JAA], 2003). This condition

is sketched in Figure 2. As the aircraft comes closer to the runway threshold, the pilot sees and aims at the designated touchdown point, normally 305 m down the runway.

The issue of RVR and HUD use

Commercial airlines are introducing HUDs to the flight deck at an increasing rate, making it beneficial to know the effect of HUD use on touchdown performance. Generalizing from earlier research (Hoerner, 1979, Martin-Emerson and Wickens, 1997), HUD use with conformal symbology should improve tracking performance during approach and landing in restricted visibility. Showing the instantaneous flight path should enable the pilot to control the aircraft better below decision height and this should reduce the variance of the touchdown point. However, there appear to be no empirical data to support such a claim. Data do exist for certification trials for HUDs currently used in civil aviation but those trials used the HUDs in a low visibility mode with guidance down to including the landing flare, in which there was no need at all for external visual cues (JAA, 2003). Such data do not export well to touchdown performance using HUD in category I flight operations where obstacle clearance below the decision height is maintained by visual maneuvering.

Allowing a too low RVR to commence approach for HUD operations will reduce approach success rate. This is not desirable from an operational standpoint. Generalizing from earlier research on visual attention and HUD there could be a risk that pilots would focus on the HUD symbology at the expense of outside cues. As the exposure time for the approach lights before decision height is between 1 and 3 seconds during minimum RVR attentional capture on the HUD symbology could actually reduce approach success rates during very low RVR conditions, despite a better tracking performance. Squeezed between operational pressure for

lower minima for HUD's presumed advantage and possible negative effects of HUD, regulators are being pressed to generate rules that are insufficiently grounded in research.

In light of the increasing use of HUDs in commercial aviation and the paucity of the existing knowledge base, we studied whether the established improvement in tracking performance when using a HUD will be reflected in a smaller touchdown footprint on the runway. We predicted that the HUD would reduce the size of the footprint both longitudinally and laterally and would not reduce the approach rate.

Experiment 1

We had initially considered conducting a series of experiments that would fully cross manipulations of RVR and the length of approach lighting to assess the interaction of these factors on HUD use during approach and landing operations. It immediately became obvious that not all combinations needed to be studied. For example, experienced commercial pilots are adept at making approaches and landings at facilities with a full system of approach lights (>720 m) when visibility is above the RVR minimum of 550 m. We also realized that the ground visual segment would be so short that a majority of pilots would opt for a missed approach rather than attempt to land when RVR was below 400 m. Accordingly, we conducted two experiments. In the first, we assessed HUD use in visibilities at and below the required minimum RVR at simulated facilities with a full system of approach lights. In the second, we assessed HUD use in visibilities at and below the required minimum RVR at simulated facilities with a system of approach lights of intermediate length (420 or 600 m).

Method

Participants. Forty-three pilots from a major European airline volunteered to participate. All were qualified to fly the B737-700 aircraft and had completed their HUD training for the operator. The training sessions consisted of one day of theory and two simulator sessions of four hours duration each. Experience on the B737-700 varied from 50 to more than 1000 hours. There is every reason to believe that these participants are representative of the population of commercial pilots to whom regulators need the data to generalize.

Apparatus. We used a CAE B737-700 training simulator with aircraft aerodynamics and visual angles valid for B737-700 to collect data. This six-axis full-motion simulator is approved for low-visibility operations down to an RVR of 200 m. The pilot had a field of vision of 180°/40° collimated to a focal distance greater than 10 m.

The HUD installed in the simulator was a Rockwell-Collins Flight Dynamics HGS-4000® (Head-up Guidance System), certified for low visibility operations down to and including RVR of 200 m. The HUD symbology used in the study is shown in Figure 2. The symbology and functionality met the production-line standard specification for the mode customarily used (IMC, instrumental meteorological condition mode) when making Category I instrument landing system (ILS) and non-precision approaches.

The HUD provided conformal display of flight path and flight-path guidance. The HUD provided the flight-path guidance using a ring inside the flight-path symbol. Flight-path guidance is not available in the conventional head-down instrumentation of the B737-NG with a primary flight display. A HUD mode used for Category 1 ILS and non-precision approaches was chosen to improve the external validity (in the form of operational usability) of the study. In

standard operational settings, the vast majority of the ILS systems are approved for only Category I operations and do not have the quality and integrity to support HUD modes incorporating landing-flare guidance.

Simulated ground facilities. The radio navigation facilities simulated in the study conformed to the International Civil Aviation Organization (ICAO, 1996) standard for ILS radio navigation aids. Such a system transmits a beam for both vertical and lateral reference. The integrity of the transmitted beam is guaranteed to keep the aircraft within allowable airspace, safe from obstacles down to the decision height. Approaches in Experiment 1 were flown to a simulated runway with a system of approach lights 900 m in length. This length falls within the full facilities system category of European aviation regulation (JAA, 2003).

Design. The experiment used a complete block design with repeated measures on HUD use (HUD, no-HUD). The order of HUD use was counterbalanced across participants to control for potential order and carryover effects. Two blocks of visibility, the standard minimum RVR (550 m) for full facilities system and lower-than-standard RVR (400 to 500 m), were a between-subjects variable. Each participant completed two approaches to a runway with a full system (900 m) of approach lights.

Procedure. Two approaches were manually flown using standard operating procedures. The scenarios started as a 6 nautical mile final to the runway in lower-than-standard or standard RVR. In the with-HUD condition, the pilots flew the aircraft by keeping the flight-path symbol centered on the guidance cue. At 50 m above the runway threshold, the guidance cue was automatically removed and the pilots performed the landing flare using external visual cues in conjunction with the HUD flight-path symbol. The automatic removal of the guidance cue and

the HGS-4000® IMC mode were deliberately chosen to insure that the pilots could not attend solely to the HUD symbology in the with-HUD trials as well as improving the external validity (in the form of operational usability) of the study. Approach and landing plots were printed using the aircraft's center of gravity as reference to determine the size of the touchdown footprint.

Results

A total of 125 approaches were flown in experiment 1. Fourteen pilots performed a go-around in one or more of approaches. Due to technical problems, data were lost from 11 pilots. The final data set consists of 43 pairs of successful approaches. Of these 43, 18 were flow in using the HUD in the first trail and 25 using the conventional head-down display (no-HUD) in the first trial; 15 were flown in standard RVR (550 m) conditions and 28 in reduced RVR conditions.

Touchdown performance. A series of full-model and reduced-model linear regressions were calculated to create separate ANOVA tables for the lateral (cross-runway) and longitudinal (down-runway) components of the touchdown footprint. The regression approach is robust to differences in sample sizes across blocks (levels of RVR) and allows for assessment of order effects and their interactions with the main factor of interest, HUD use. We adopted an alpha level of .05 for statistical significance.

Lateral touchdown performance was measured as the absolute lateral deviation from the runway centerline at touchdown. The data for the main effect of HUD use is shown in Figure 4a. There was a strong effect for HUD use, $F(1, 78) = 8.991$, $p < .004$, $\eta^2 = .12$ at an alpha of .05 indicating a power of at least .71. Landings were closer to the centerline when pilots used the

HUD. The main effects for RVR and order of HUD condition were not significant. There was, however, a marginally significant interaction between RVR and the order of HUD use, $F(1, 78) = 3.235$, $p < .076$, $\eta^2 = .11$. As shown in Figure 4b, both the group that used the HUD on their first approach (HUD 1st) and the group that used the HUD on their second approach (no HUD 1st) landed nearer the centerline when they used the HUD. However, the improvement is greater for the group that used the HUD in the second approach.

Longitudinal touchdown performance was measured as the distance from the runway threshold at touchdown. Contrary to expectations and in contrast with the results for lateral performance, the effect of HUD use was not significant. However, the between-subjects main effect of RVR, shown in Figure 5a, was significant, $F(1, 78) = 6.504$, $p < .013$, η^2 .13 indicating a power of at least .65. The mixed-effect interaction of RVR (between subjects) and HUD sequence (within subjects) was also significant, $F(1, 78) = 5.750$, $p < .019$, Figure 5b. Post-hoc analysis suggests that landings were considerably shorter in low RVR for the group that used the HUD first.

Approach success rate. Approach success rate was defined as the number of registered successful approaches, divided by the total number of registered approaches. A chi-square test was used to examine the independence of landing success and HUD use. The observed data from all 125 approaches, collapsed across levels of RVR, are shown in Table 2. The test indicates the two factors are, as expected, independent, $\chi^2 (1) = 1.36$, $p > .24$. Similar results were found within the low and standard manipulations of RVR.

Experiment 2

Method

The method, procedure, and design used in the second experiment were identical to those used in the first experiment with the few exceptions discussed here. The different criteria for standard RVR across facility types preclude collapsing and analyzing the data the data as a single experiment.

Participants. Twenty-eight pilots from the same major European airline volunteered to participate. None of the volunteers had participated in Experiment 1. All were qualified to fly the B737-700 aircraft and had completed their HUD training for the operator. Experience on the B737-700 varied from more than 50 to more than 1000 hours.

Simulated ground facilities. Approaches in Experiment 2 were flown to simulations of the runways at Bruxelles (ILS 25R BRU) with 600 m of approach lights or at Tromsö (ILS 01 TOS) with 420 m of approach lights. These lengths fall within the intermediate facilities category of approach lights as defined by European aviation regulation (JAA, 2003).

Design. Two blocks of visibility, the standard minimum RVR (700 m) for intermediate facilities and lower-than-standard minimum RVR (550-600 m), were a between-subjects variable in a complete-block repeated-measures design. Each participant completed at least two approaches to a runway with a system of approach lights either 420 or 600 m in length. Nine of the participants completed more than two approaches each but only the first in each condition is used in the analyses of the size of the touchdown footprint.

Results

A total of 128 approaches were flown in experiment 2. On 17 approaches the pilots performed a go-around in one or more approaches. Due to technical problems data were lost from 6 approaches. The final data set consists of 28 pairs of successful approaches. Of these 28, 14 were flow in using the HUD in the first trail and 14 using the conventional head-down display (no-HUD) in the first trial; 14 were flown in standard RVR (700 m) conditions and 14 in reduced RVR conditions.

Touchdown performance. No differences were found across the two intermediate lengths of runway lights (420 m vs. 600 m). The analyses presented here collapse the data across this variable.

The main effect for HUD use, shown in Figure 5, was the only factor in the ANOVA to approach significance. As in Experiment 1, the effect for HUD use on the lateral component of the touchdown footprint was strong, $F(1, 48) = 7.380$, $p < .009$, $\eta^2 = .06$ indicating a power of at least .54. Once again, landings were closer to the centerline when pilots used the HUD.

As in Experiment 1, there was no significant effect for HUD use on the longitudinal component of the touchdown footprint. Unlike Experiment, 1 there was also no effect of RVR condition.

Approach success. Table 3 shows the frequencies (n = 128) for the intermediate approach lights facilities (length 420 - 600 m). Approach success rate was 0.86 for HUD and 0.88 without HUD. A chi-square test was used to examine the independence of landing success and HUD use. As in

Experiment 1, the test indicates the two factors are, as expected, independent, χ^2 (1) = 0.053, p > .81. Similar results were found within the low and standard manipulations of RVR.

Discussion

There are three findings. First, HUD use did not influence the pilots' decision making at the decision height regarding landing or making a go-around. The lack of an effect suggests that the additional information made available by the HUD did not distract the pilots' attention or interfere with their decision making process during the most critical portion of the approach and landing sequence. Second, HUD use significantly reduced variability in the lateral component of the touchdown footprint for all RVR conditions. In contrast, it appears that the HUD does not influence the size of the longitudinal footprint. The first two findings conform to our hypotheses. Here we reexamine our hypotheses about the impact of HUD use on the touchdown footprint and offer an explanation for its differential impact on the lateral and longitudinal components.

Experience engenders automaticity. For commercial pilots, hours of practice and numerous repetitions make touchdowns at standard facilities in good visibility and weather relatively routine. Nevertheless, the ubiquitous and ever-varying direction and velocity of crosswinds are likely to preclude the development of true automaticity at touchdown. Crosswinds introduce an element of uncertainty regarding drift (the lateral location of the aircraft relative to the centerline). For the pilot to detect any drift the visual ground segment needs to be long enough determine the aircraft's movement versus the ground. That means that to detect drift at all, a noticeable lateral displacement must take place and not all of this displacement can

be identified before touchdown. Such exogenous uncertainty is likely to have conditioned pilots (and regulators) to tolerate a certain amount of variance in the lateral touchdown footprint.

The HUD largely eliminates uncertainty about drift. The addition of a conformal flight-path vector provides instantaneous feedback about aircraft drift and actual flight path. The additional information enables precise control of the flight path during approach and landing and reduces the variance in lateral displacement practically to nil.

In contrast, the HUD mode examined provides no flare guidance and no additional information than can be used to guide the decision of when to initiate a landing flare. A small change in the timing of a landing flare at the nominal glide slope of 3° results in large longitudinal differences. Control of the longitudinal component of the touchdown footprint is largely an effect of how pilots handle the aircraft's energy (its sink rate and speed) in the final seconds before landing. Information used to control the aircraft's energy is provided by the optic flow from the looming runway (Lee, 1974) and, in large commercial air transport aircraft, aided by a radio altimeter that counts down from 50 feet to 0 feet (runway contact) in 10 foot increments. The initiation of the landing flare has been shown to be a function of time-to-contact (Mulder, Pleijsant, van der Vaart, & van Wieringen, 2000). It is easy to understand, then, why HUD use had no effect on longitudinal touchdown performance. It remains to be seen whether similar results are found for HUD modes with flare guidance, and which visual representations are actually most effective in prompting pilots to reduce sink rate and speed.

Our data suggest that low RVR is likely to influence the point of touchdown.. In Experiment 1, pilots consistently landed shorter in low RVR conditions than in standard RVR conditions. We attribute this observation to a change in the visual horizon. In low RVR, the

short visual segment shifts the visual horizon closer to the runway threshold. If the pilot uses the visual horizon as a cue to initiate the landing flare, the landing flare is likely to be initiated late. A late flare reduces the time the aircraft glides above the runway and leads to a relatively short touchdown points, like those observed in our experiment We attribute the interaction of HUD sequence and RVR observed in Experiment 1 to the focus of attention. During approach in low RVR, pilots follow their instruments; they focus their attention on the HUD symbology. It is likely that pilots focusing their attention on HUD symbology assimilate outside cues later than pilots who have no HUD guidance.

The effects of HUD use on landing performance were seen in standard and lower-than-standard RVR conditions in both Experiment 1 and 2. This carries a clear and important operational implication: Based on empirical performance data, the minimum RVR for approaches using HUD could be set lower than for approaches without a HUD while still maintaining a similar or even smaller touchdown footprint.

Conclusions

Data from the experiments reported here show that HUD guided operations improve lateral touchdown performance, likely because the conformal flight path vector makes it easier for pilots to determine, and correct for, aircraft drift. We did not find an effect of HUD use on longitudinal touchdown performance, probably because of the lack of flare guidance in the HUD modes studied here. However, the other beneficial effects of HUD use on landing performance were seen in standard and lower-than-standard RVR conditions in both experiments, which implies that the minimum RVR for approaches using HUD could be set lower than for approaches without a HUD. Finally, we did not find any detrimental effect of HUD use on approach success rate.

These results could be critical in supporting regulatory decision making about the visibility required to commence an approach using a HUD-equipped aircraft. In fact, it is precisely this type of systematic empirical data that is often lacking as a basis for developing well-informed regulations, in Europe as well as the USA. We believe that empirical results like these should be a critical ingredient in safety regulators' rule-writing for (the application of) new technology in commercial aircraft operations. When it comes to HUD, the need for informed regulation of HUD use is already pressing, and will only increase as HUD use proliferates. The relevance of the findings here could even support the expansion of such targeted empirical studies, for example to Category II conditions, or to refine them to, for example, find the lowest acceptable RVR using HUD.

References

Fadden, S., Ververs, P. M. &Wickens, C. D. (1998). Costs and Benefits of head-up display use: A meta-analytic approach. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*,16-20, Santa Monica, CA.

Flach, J. M. &Rasmussen, J. (2004). Cognitive Engineering: Designing for situation awareness. In Sarter, Amalbert (Ed.), *Cognitive Engineering in the Aviation Domain* (pp. 153-180). Mahwah, N.J: Lawrence Erlbaum.

Flight Safety Foundation. (1991). Special Safety Report: Head-up Guidance Technology (HGST). A Powerful Tool for Accident prevention. *Flight Safety Digest*, FSF/SP-91/91, Alexandria, VA.

Gibson, J. J. (1979, 1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates:

Iavecchia, J. H., Iavechia, H. P., & Roscoe, S. N. (1988). Eye accommodation to Head-Up virtual images. *Human Factors 30*, 689-702.

ICAO. (1996). *Annex 10, Aeronautical Telecommunication, Volume I – Radio Navigation, 5th edition*. Montreal, Quebec: International Civil Aviation Organization.

ICAO. (2001). *Annex 6,Operation of Aircraft, 8th edition*. Montreal, Quebec: International Civil Aviation Organization.

ICAO. (2003). *OPS – Aircraft Operation (Doc 8168) Volume II, 4th edition*. Montreal, Quebec: International Civil Aviation Organization.

Joint Aviation Authorities. (2003). *All Weather Operations JAR-OPS 1 Subpart E*. Hoofddorp, The Netherlands.

Lauber, J. K., Bray, R. S., Harrison, R. L., Hemingway, J. C., & Scott, B. C. (1982). An operational evaluation of head-up displays for civil transport aircraft. *NASA Tech. Paper 1815*. Moffett Field, CA: NASA Ames Research Center.

Lee, D. (1976). A theory of visual control of breaking based on information about time to collision. *Perception 5*, 437-459.

Martin-Emerson, R. & Wickens, C. D. (1997). Superimposition, symbology, visual attention and the head-up display. *Human Factors 39*(4), 581-601.

Mulder, M., Pleijsant, J-M., van der Vaart, H., & van Wieringen, P. (2000). The effects of pictorial detail on the timing of the landing flare: Results of a visual simulation experiment. *The International Journal of Aviation Psychology, 10*(3), 291-315.

Newman, R. L. (1995). *Head-up displays: Designing the way ahead*. Aldershot: Ashgate.

Sanford, B. D., Foyle, D. C., McCann, R. S. & Jordan, K. (1993). Head-Up Displays: Effect of Information Location on the Processing of Superimposed Symbology. *Proceedings of the 7th International Symposium on Aviation Psychology*, pp. 81-87. Columbus, OH

Stuart, S. W., McAnally, K. I. &Meehan, J. W. (2003). Head-up displays and visual attention: Integrating data and theory. *Human Factors and Aerospace Safety 1*(2), 103-124

Weintraub, D. J, Haines, R. F. & Randle, R. J. (1985). Head-up display (HUD) utility, II: Runway to HUD transitions monitoring eye-focus and decision times. *Proceedings of the*

Human Factors Society 29th Annual Meeting, (pp. 615-619). Santa Monica, Ca: Human Factors Society

Wickens, C.D. & Long, J., (1995). Object versus space-based models of visual attention: Implications for the design of Head-Up displays. *Journal of Experimental Psychology: Applied 1*(3), 179-193.

Wolffsohn, J. S, Edgar, G. K., Stone, H. E., Williams, M. & Mc Brien, N. A. (1999). Does Over-Accommodation Occur When Using Aircraft Head-Up Displays? *Aviation, Space and Environmental Medicine, 70*(7), 666-673.

Author notes

Correspondence should be addressed to Örjan Goteman, Flight Operations Standards, STOOS, Scandinavian Airlines, S-195 87 Stockholm, Sweden; orjan.goteman@sas.se

The authors thank Staffan Carlsson and Zacharias Soudah for assistance in data collection, Jan-Olof Ehk for statistical advice, and the pilots who volunteered their time and skill. Rockwell Collins Flight Dynamics Company provided simulator time to conduct the experiment.

Örjan Goteman is a captain for Scandinavian Airlines flying B737. He is also responsible for specifying and implementing new aviation systems across the entire SAS aircraft fleet. He received a university medical degree from the University of Linköping in 1987 and is on schedule to receive his Ph.D. in Industrial Ergonomics from the Linköping Institute of Technology in 2005.

Kip Smith is a research professor of Industrial Ergonomics at the Linköping Institute of Technology, Sweden. He received his Ph.D. in management with an emphasis in decision science from the University of Minnesota in 1996.

Sidney Dekker is Professor of Industrial Ergonomics at the Linköping Institute of Technology, Sweden, and is Director of Studies at the Centre for Human Factors in Aviation there. He gained his Ph.D. in Cognitive Systems Engineering from The Ohio State University, USA.

Tables

Table 1. Runway visual range (RVR) standards for facilities with full and intermediate systems of approach lights.

Experiment 1 with Full Facilities (900 m)		Experiment 2 with Intermediate Facilities (420-600 m)	
Low RVR	Standard RVR	Low RVR	Standard RVR
400-500 m	550 m	450-650 m	700 m

Table 2. Contingency table of the pilots' decisions to land or go-around across conditions of HUD use at simulated facilities with a full system of runway lights (900 m), Experiment 1.

	Go around	Landing	
HUD	9	53	62
No HUD	5	58	63
	14	111	125

Table 3. Contingency table of the pilots' decisions to land or go-around across conditions of HUD use at simulated facilities with an intermediate system of runway lights (420 or 600 m), Experiment 2.

	Go around	Landing	
HUD	10	62	72
No HUD	7	149	56
	17	111	128

Figure captions

Figure 1. Typical HUD symbology with Flight Path symbol as a ring with slanted wings in the middle of the display. The horizontal line in the middle of the display represents the horizon. The aircraft symbol at the top of the display represents how the aircraft's longitudinal axis is pointing, in this case seven degrees up and slightly to the left of the actual track. To the left is the speed gage in the form of a speed tape and to the right is the altimeter in the form of an altitude tape. Present speed is 173 knots and altitude is 1980 feet. The relative location of the flight path symbol on the horizon line indicate that, in this example that the aircraft maintains its altitude. (Courtesy of Rockwell-Collins Flight Dynamics).

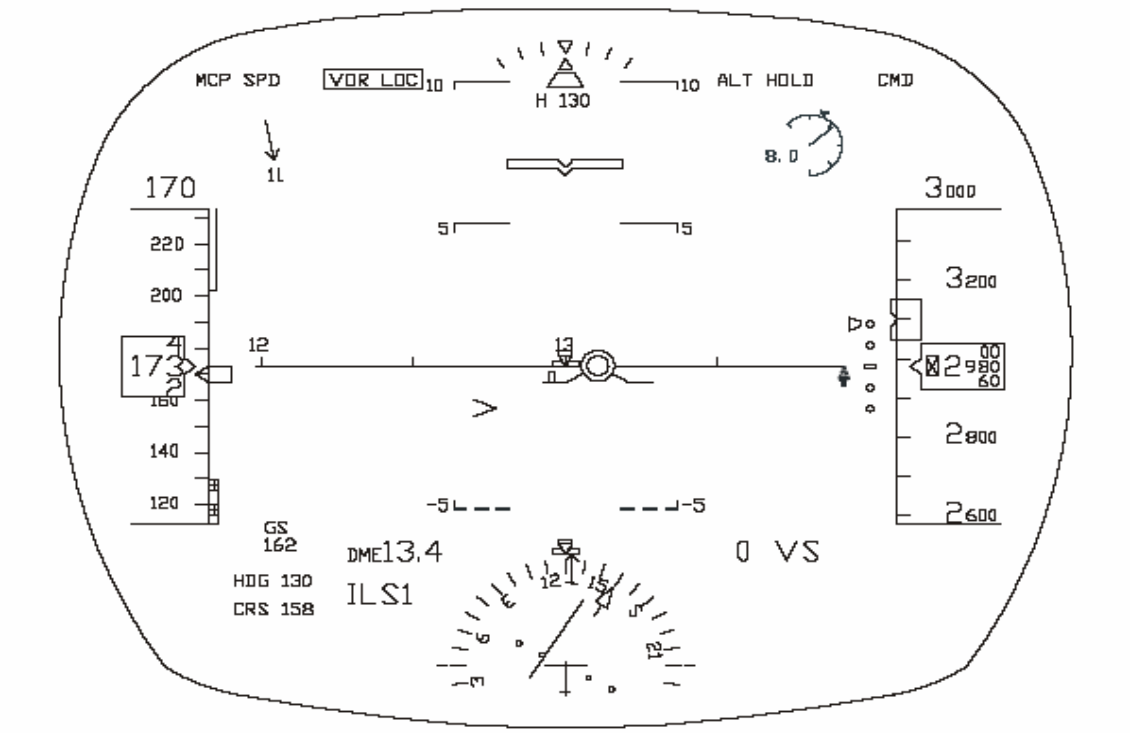
Figure 2. The visual segment as seen from the aircraft at decision height at a given the runway visual range (RVR).

Figure 3. The HUD symbology in IMC mode used during the approaches in Experiments 1 and 2.

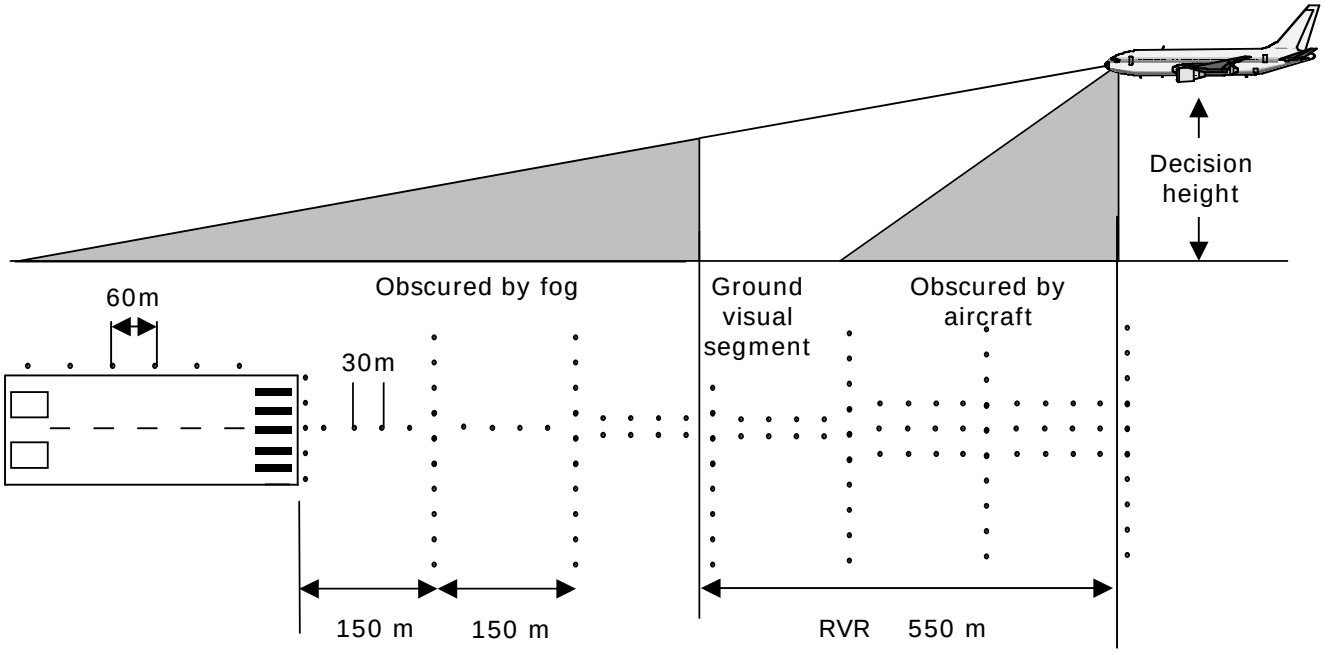
Figure 4. Graphs showing (A) the main effect for HUD use and (B) the interaction of HUD use and the order of its use on the lateral size of the touchdown footprint at simulated facilities with a full system (900 m) of approach lights, Experiment 1.

Figure 5. Graphs showing (A) the main effect for RVR and (B) the interaction of RVR and trial order of HUD use on the longitudinal size of the touchdown footprint at simulated facilities with a full system (900 m) of approach lights, Experiment 1.

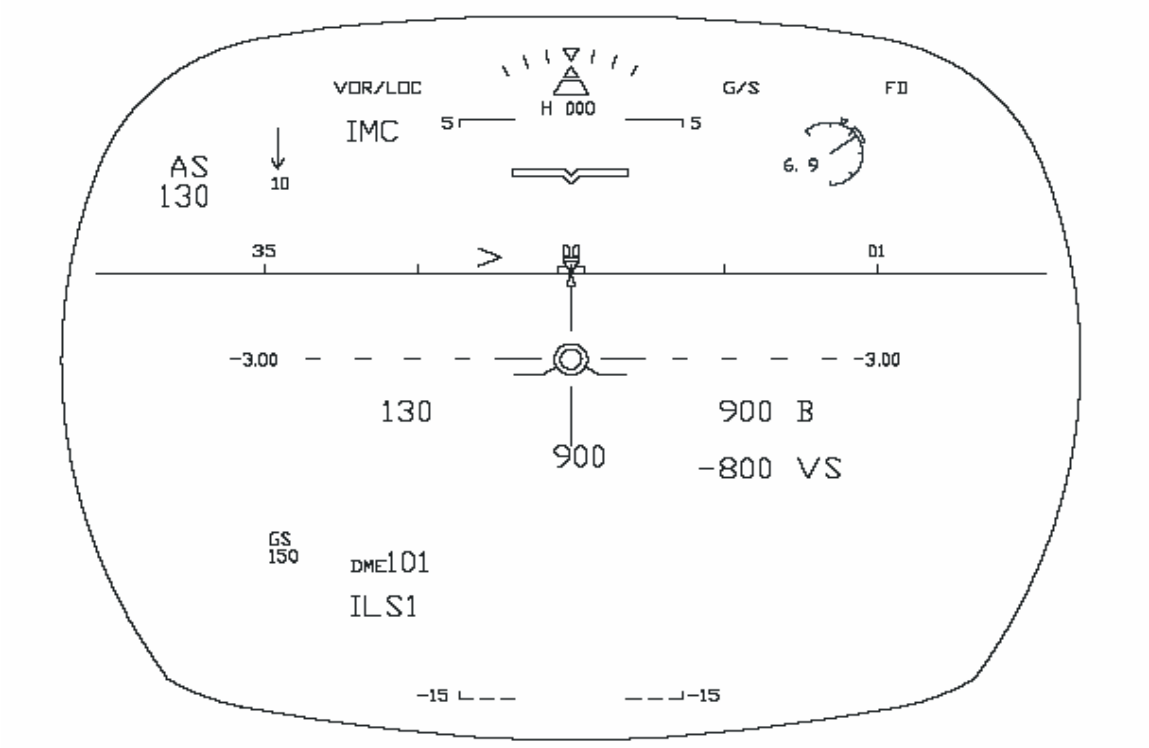
Figure 6. Graph showing the main effect of HUD use on the lateral size of the touchdown footprint at simulated facilities with an intermediate system (420 m or 600 m) of approach lights, Experiment 2.



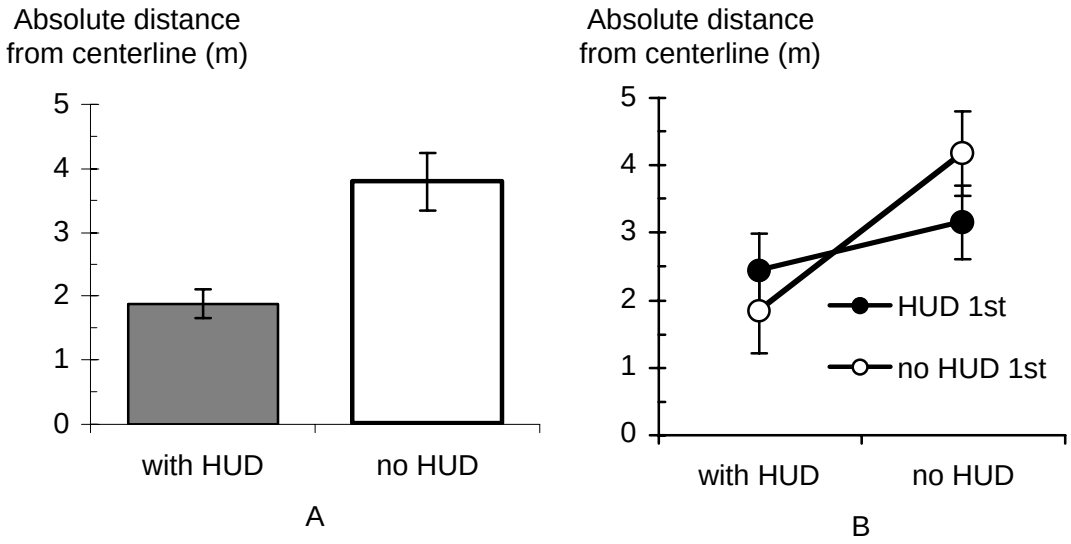
HUD use Figure 1



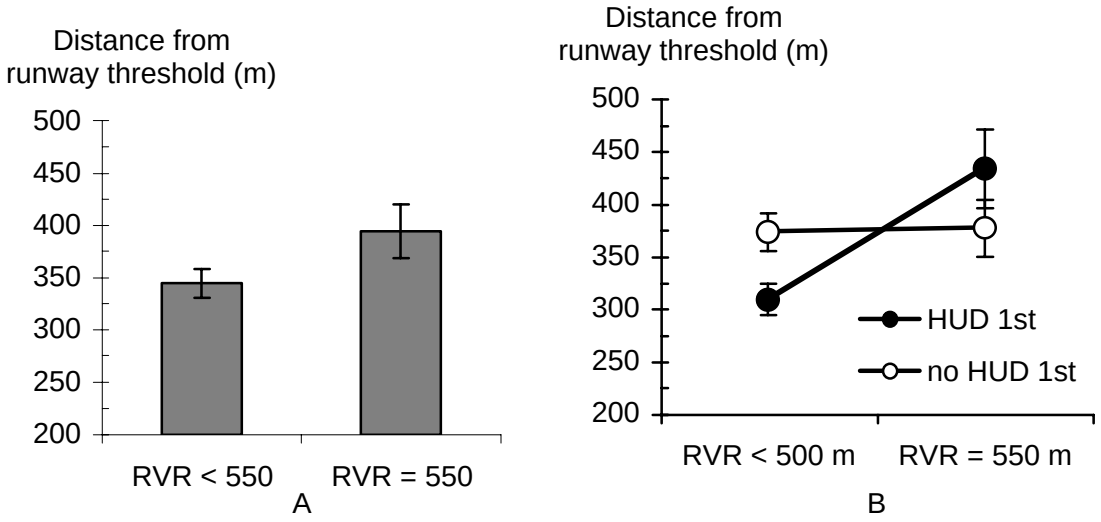
HUD use Figure 2



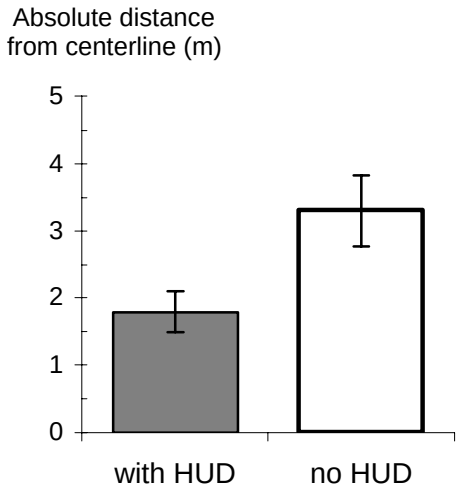
HUD use Figure 3



HUD use Figure 4



HUD use Figure 5



HUD use Figure 6

Linköping University Dnr IKP 214/03-23
Progress Report covering the period 1 April 2003 - 31 June 2003
MURI OPUS -- DAAD19-01-1-0621

Principal Investigator: Kip Smith, Ph.D.
Division of Industrial Ergonomics

This progress report reviews work conducted during the second quarter of 2004 for the MURI OPUS research program at Linköping University (LiU), Sweden. During the second quarter, LiU received an unexpectedly large (18%) cut in MURI OPUS funding. As a result, we have had to cut our research effort significantly. Nevertheless, we conducted the fourth in our series of paintball experiments and wrote drafts of a journal article on the first three experiments and of a proceedings paper on the fourth experiment.

The mission of the LiU research program has been and continues to be the identification and mitigation of cognitive decrements due to remote command and control under realistic simulations of battlefield conditions (on the paintball arena). The first half of that mission – documenting the existence of a decrement in a soldier's response time when the fire-team leader is not present in the field with the soldier – was successfully completed during the first quarter of 2004. The experiments completed during the first quarter document a cognitive decrement directly attributable to remote command and control. The attached document is the draft journal article on these experiments. The target journal is Organizational Behavior and Human Decision Processes.

The experiment conducted in the second quarter of 2004 is the first in a new series that investigate a mechanism for mitigating the identified cognitive decrement. The success of the proposed mitigation has significant implications for the future of the dismounted infantry and for the design of remote command-and-control systems generally.

The first section of this report reviews how we met the goals for the second quarter that were outlined in the progress report for the first quarter. The second section introduces the goal for the second half of 2004. Attached is a draft article on the experiment conducted in the second quarter that being circulated to reviewers prior to submission to the Journal of Experimental Psychology – General.

Both research goals have been met

The first quarter report itemized two goals for the second quarter. Both have been met.

1 Solidify and extend research links with Försvarshögskolan (the Swedish defense college), Högskolan Skövde (the University at Skövde), and the garrison at Skövde.

Rogier Woltjer, a Ph.D. student from LiU, is conducting research on command and control operations at Försvarshögskolan under Professor Smith's supervision. The project is supported by the Försvaretsmaterielverket, the

Swedish defense material administration. The paintball experiments are being conducted with the assistance of students and faculty from Högskolan Skövde. Skövde is the home of four regiments of the Swedish Army. The Colonel who is the head of the Skövde garrison is keen to provide access to recruits as participants in our experiments.

2 Submit for publication the article on leader presence.

The attached document is the draft of the article. Comments are welcome. Please email them to Prof. Smith at kipsm@ikp.liu.se

Goal for the third and fourth quarters of 2004:

Conduct an experiment that replicates and extends the four previous experiment in two critical ways:

- Use a between-subjects (rather than within-subjects) design with all four of the conditions used in the previous experiments. The four conditions are numbered in the matrix below. The between-subjects design will eliminate the possibility of carry-over effects.

Communication	Leader Presence	
	Collocated	Remote
Face-to-face	1	-
Mediated by radio	2	3
Mediated by ratio and tactile stimuli	-	4

- Use recruits from the garrison at Skövde as participants. Data from soldiers-in-training rather than college students are more likely to persuade both the U. S. and Swedish Armies to adopt a method of tactile stimuli as they develop plans for remote command-and-control operations.

Linköping University Dnr IKP 214/03-23
Progress Report covering the period 1 October 2004 - 31 December 2004
MURI OPUS -- DAAD19-01-1-0621

Principal Investigator: Kip Smith, Ph.D.
Division of Industrial Ergonomics

Progress

During the fourth quarter of 2004, three articles were submitted for publication to peer-reviewed journals and one abstract submitted to an international conference. The list of MURI OPUS - supported publications for the year 2004 is attached.

No new research was conducted during the quarter because the entire LiU budget had been exhausted by the end of the third quarter. Of the three factors responsible for this situation, only one was under our control. That factor was the decision to take full advantage of a unique opportunity to collect data during the first and second quarters. The two paintball experiments conducted during the first half of the year largely exhausted the limited budget. Travel to present MURI-supported research consumed the remaining funds.

Economic reality

The two factors beyond our control that contributed to the early expenditure of all grant funds was the 20% decline in the value of the dollar relative to the Swedish crown (and all other currencies) and the unexpectedly harsh 17% cut in our budget. [Details. When I moved to Sweden in the summer of 2003, the exchange rate was 8.5 crowns/dollar. The rate is now (December 2004) 6.6 crowns/dollar. The original budget for 2004 was approximately $90,000. The revised budget $75,000.] This combination effectively reduced the time available to conduct research by more than a third. A similarly dire situation is forecast for the remainder of the project.

Goal

Nevertheless, our long-term goal continues to be the identification and mitigation of cognitive decrements in soldier performance due to remote command and control. To this end, we plan to conduct additional experiment on leader presence under live paintball fire during the first half of 2005. We hope to conduct longitudinal experiments at the military garrison in Skövde, Sweden, with active-duty soldiers and officers as participants. These experiments will replicate and extend those discussed in the articles that have been submitted for publication. Among the issues we hope to investigate are the relative effects of leader presence and prior history.

Goals for the first half of 2005:

* Conduct research at the garrison at Skövde with assistance from faculty and students at Högskolan i Skövde.

* Submit for publication additional articles on leader presence.

MURI – supported publications by Kip Smith, Linköping Institute of Technology

Articles in review:

Smith, K., Liberg, M., Hilmersson, J. & Sitzes, R. (2004). Simulated touch: A method for stimulating trust. Manuscript submitted for publication.

Smith, K., & Pangburn, K. (2004). The value of leader presence. Manuscript submitted for publication.

Smith, K. & Hancock, P. A. (2004). Time to contact in representational space. Manuscript submitted for publication.

Goteman, Ö., Smith, K., & Dekker, S. (2004). HUD use improves landing performance in restricted visibility. Manuscript submitted for publication.

Articles accepted but not yet published:

Knecht, W. R. & Smith, K. (In press). The maneuver space: Theoretic lessons learned from empirical practice. International Journal of Aviation Psychology.

Smith, K. (In press). The adverse impact of remote command and control under live fire. Human Performance in Extreme Environments.

Article published in 2004

Murphy, L. L., Smith, K., Hancock, P. A. (2004). Task demand and response error in a simulated air traffic control task: Implications for ab initio training. International Journal of Applied Aviation Studies, 4(1), 91-106.

Edited book published in 2004

Smith, K., Shanteau, J., & Johnson, P. E. (Eds.). (2004). Psychological Explorations of Competent Decision Making. Cambridge, UK: Cambridge University Press.

Book chapter published in 2004

Grazioli, S., Smith, K., & Johnson, P. J. (2004). Managing risk in social exchange. In K. Smith, J. Shanteau, & P. E. Johnson (Eds.), Psychological Explorations of Competent Decision Making (pp. 71-123). Cambridge, UK: Cambridge University Press.

Conference papers and presentations in 2004

Woltjer, R. & Smith, K. (2004). Decision support through constraint propagation in collaborative distributed command and control. Proceedings of the 23rd European Annual Conference on Human Decision Making and Manual Control. Delft, NL: Delft University of Technology.

Goteman, Ö., Smith, K., & Dekker, S. (2004). HUD use improves landing performance in a full-motion simulator. Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, 16 – 20.

Smith, K. (2004). Initial experiments on leader presence and communication mode on combat performance. Proceedings of the Second Conference on Human Performance, Situation Awareness, and Automation Technology, 2, 266-271.

Submissions to conferences in 2005

Smith, K. (2005). The structure of distributed command and control. Submitted to the Proceedings of the Seventh International Conference on Naturalistic Decision Making.

Smith, K. & Liberg, M. (2005). Simulated touch stimulates trust in a remote leader. Submitted to the Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting.

The Effects of Event Rate and Time on Task on Human Performance and Subjective Ratings During a Multi-task Infantry Simulation

Marla Zinni (1), Rebecca Proto (1), Wayne Harris (2), & Raja Parasuraman (1)
The Catholic University, Washington, DC (1) and The Human Factors Research Lab, University of Minnesota (2)

INTRODUCTION

- With the U. S. military converting to a more information based warfare concern has increased in the development of displays to show data to platoon leaders and soldiers.
- There is specific interest in the effect this will have on human performance in high workload conditions, in which the operator is fatigued .

THEORIES OF STRESS, WORKLOAD, AND PERFORMANCE

- Two level compensatory control model (Hockey, 1997)
  Decision between achieving goals for a given task and conserving energy
  3 modes of coping:
  - Active coping - increase of effort and energy use that does not result in a considerable drain on energy stores;
  - Strain coping - high use of energy to maintain task goals causing levels of anxiety and fatigue at a level that could potentially become problematic if maintained over a long period of time;
  - Passive coping - decision is made in favor of energy conservation at the cost of decreased task performance, which could take the form of a decrease in accuracy, speed, or secondary task performance.
- Adaptive control of mental workload (Parasuraman & Hancock, 2001)
  There is not necessarily a direct association between task load imposed from external environmental sources and mental workload. Workload is also mediated by the individual operator's personal characteristics (i.e., skill level, task management strategies).
  Adaptation may take the form of reduced processing, task prioritizing, task shedding or temporary allocation to automated control.

RESEARCH OBJECTIVE

- Observe the effects of task difficulty and time on task both in performance and subjective ratings of stress and workload.

Experimental Hypotheses:

- Operators would not be able to maintain their performance under conditions of high workload for extended periods of time.
- When operators are exposed to high workload for short periods of time, even when maintaining performance, energetic costs in terms of subjective ratings would still be incurred.

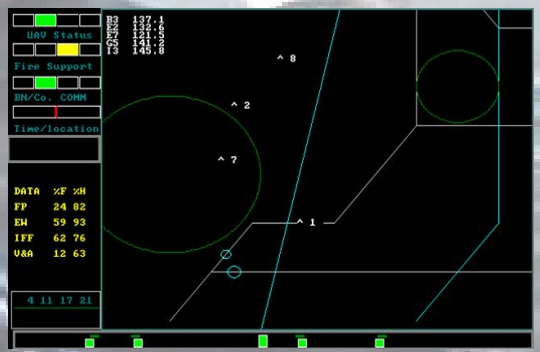
METHODS

- Participants
  - 15 young adults
  - 18-27 years (M = 21.56 years, SD = 3.16 years)
- Design 2x2 within subjects
  - 2 Difficulty Levels (Hard and Easy)
  - 2 Task Durations (10 minutes and 30 minutes)
- Dependent Variables
  - Decision Accuracy
  - Response Time
  - Subjective Ratings (stress and workload)

VIKING ENVIRONMENT

Viking Infantry Information and Communication Environment
-(Harris, Parasuraman, Zinni, Hancock, Harris, 2002)

- 4 tasks:
  1) plane classification task
  2) link reset task
  3) light monitoring task
  4) platoon position and pace scheduling task



RESULTS

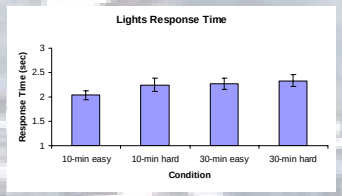
TASK PERFORMANCE

Dependent Variables	10-min easy	10-min hard	30-min easy	30-min hard
Plane Accuracy (% correct)	88.00	88.33	84.22	89.96
Plane Response Time (sec)	3.03	3.08	3.33	3.11
Light Accuracy (% correct)	98.67	94.22	97.48	93.70
Light Response Time (sec)	2.04	2.25	2.27	2.33
Link Accuracy (% correct)	99.67	98.00	99.33	98.00
Link Response Time (sec)	2.19	2.38	1.98	2.20
Platoon Position (unit deviation from ideal)	5.31	6.39	7.48	7.89
Platoon Pace (units behind schedule)	-4.27	-2.77	-14.39	-18.68

The performance data revealed decreased performance in some of the tasks and maintenance in others. The following variables showed significant decreases in performance for time on task: light monitoring response time (F (1,14) = 4.641, p < .05), link resetting response time (F (1,14) = 8.501, p < .05), and platoon pace (F (1,14) = 5.956, p < .05). Platoon position also showed a decreased performance trend that almost reached significance (F (1,14) = 3.785, p = .07). Link response time (F (1,14) = 18.428, p < .01) and link accuracy (F (1,14) = 5.560, p < .05) showed significant decreases in performance for task difficulty.

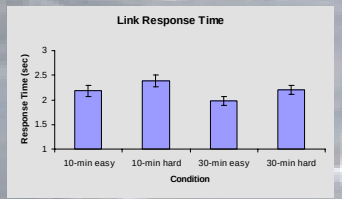
LIGHT MONITORING

The light monitoring task consisted of responding to a visual cue simulating an incoming message warning. Performance on this task significantly decreased for time on task.



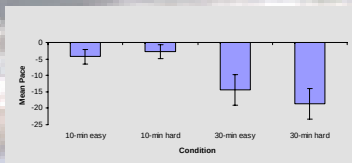
LINK RESETTING

Link resetting consisted of a number monitoring task simulating communication link failure. Performance on this task significantly decreased for both time on task and difficulty level.



PACE SCHEDULING

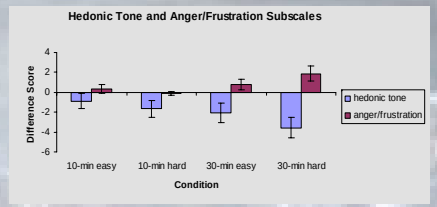
Pace scheduling refers to the need of maintaining a desired platoon pace. Performance on the pace scheduling task decreased significantly with time on task, i.e. platoons are further behind schedule for the 30 minute conditions.



DSSQ RESULTS

(Dundee Stress State Questionnaire)

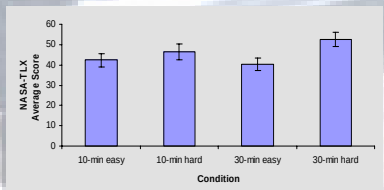
Participants showed a more negative mood state during the longer sessions, F (1, 7) = 9.398, p < .05, as reflected by a significant decrease in hedonic tone.
A significant interaction was observed between task difficulty and time on task for the 'anger and frustration' subscale, F (1,8) = 9.561, p < .05. With the highest scores indicating more 'anger and frustration' for the 30 minute hard condition.



NASA-TLX RESULTS

(NASA-Task Load Index)

A significant main effect was observed for task difficulty, as reflected on the NASA-TLX average score, F (1,14) = 9.521, p < .01. Higher levels of task difficulty resulted in higher subjective ratings of mental workload.
There was also a significant interaction between difficulty and time on task, F (1,14) = 4.626, p < .05, with the highest NASA-TLX score for the 30 minute hard condition.



CONCLUSIONS

The decreased performance on simulated tasks such as incoming message warning, communication link failure, and platoon pacing and positioning, reflects the difficulties in maintaining performance on situations that should occur in a real world scenario, supporting the use of the Viking task in human performance studies.

Our data revealed partial maintenance of performance. That is, we only found significant decreases in performance for some of the assessed variables.

Regarding energetical costs, our results revealed an increase in workload with indications of concomitant increases in stress and frustration.

Together, our study revealed that with increased difficulty and prolonged time on task, participants showed a partial decrease of performance accompanied by an increase of energetic costs. These findings bring support to both hypotheses proposed. Moreover, they are in agreement with the idea of 'strain coping', as defined by Hockey (1997), in his compensatory control model of human performance.

ACKNOWLEDGEMENT

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

REFERENCES

Harris, W. C., Parasuraman, R., Zinni, M. R., Hancock, P. A. & Harris, S. C. (March, 2002). *Viking: Development of an infantry scenario multi-task environment.* American Psychological Association Division 21 Midyear Symposium, Fort Belvoir, VA.
Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 35*, 73-93.
Parasuraman, R. & Hancock, P. A. (2001). Adaptive Control of Mental Workload. In P. A. Hancock & P. A. Desmond (Eds.) *Stress, Workload, and Fatigue*, New Jersey: Lawrence Erlbaum Associates.

STANDARD

NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION

AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

AGREEMENT TITLE: Operator Under Stress Research

AGREEMENT NUMBER: NCRADA-**NAVAIR ORLANDO TSD-02-037**

AGREEMENT ADMINISTRATORS:

NAVAIR ORLANDO TSD

Technology Transfer Office:	**Ms. Patricia R. Nimmo, AIR 49T, (407) 380-8225**
Legal Counsel:	**Robert W. Adams, AIR 772P, (407) 380-8211**
Principal Investigator:	**Ron Wolff, AIR 4962, (407) 380-4583**

UCF MURI / OPUS

Preferred Contact:	**Mrs. Kim Smith, UCF Office of Research, (407) 823-3062**
Principal Investigator:	**Dr. Peter Hancock, Director, UCF MURI/OPUS Laboratory, (407) 823-2310**

This Page Intentionally Left Blank

TABLE OF CONTENTS

PREAMBLE

Article 1.
DEFINITIONS 2

1.1 "Agreement" 2
1.2 "Classified Information" 2
1.3 "Collaborator" 2
1.4 "Controlled Unclassified Information (CUI)" 2
1.5 "Cooperative Work" 2
1.6 "Data" 2
1.7 "Effective Date" 2
1.8 "Exclusive License" 2
1.9 "Government" 2
1.10 "Government Purpose Rights" 2
1.11 "Information" 2
1.12 "Intellectual Property" 2
1.13 "Invention" 2
1.14 "Invention Disclosure" 3
1.15 "Made" 3
1.16 "Militarily Critical Technologies (MCT)" 3
1.17 "Non-Subject Data" 3
1.18 "Non-Subject Invention" 3
1.19 "Patent Application" 3
1.20 "Principal Investigator (PI)" 3
1.21 "Proprietary Information" 3
1.22 "Restricted Access Information" 3
1.23 "Subject Data" 3
1.24 "Subject Invention" 3
1.25 "Tangible Property" 4
1.26 "Unlimited Rights" 4

Article 2. OBJECTIVES 4

Article 3. RESPONSIBILITIES 4

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities 4
3.2 **UCF MURI / OPUS Laboratory** Personnel and facilities 5
3.3 Security Regulations and Directives 5

Article 4. REPRESENTATIONS AND WARRANTIES 5

4.1 **NAVAIR ORLANDO TSD** Representations and Warranties 5
4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Representations and Warranties 5
4.3 Joint Representations 6

Article 5. FUNDING 7

5.1 Payment Schedule..... 7
5.2 Insufficient and Excess Funds..... 7
5.3 No New Commitments..... 8
5.4 Accounting Records..... 8

Article 6. REPORTS AND PUBLICATIONS..... 8

6.1 Final Reports..... 8
6.2 Agreement to Confer Prior to Publication or Public Disclosure..... 9
6.3 Classified Information 9

Article 7. INTELLECTUAL PROPERTY..... 9

7.1 Data..... 9
7.1.1 General Provisions Applying to All Data..... 9
7.1.1.1 Ownership..... 9
7.1.1.2 No Implied License..... 9
7.1.1.3 Marking of Data..... 10
7.1.1.3.1 Data Provided With Less Than Unlimited Rights.. 10
7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information..... 10
7.1.1.3.3 Data That are Subject to 35 USC 205..... 11
7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted..... 11
7.1.1.4 Protection of Data..... 11
7.1.1.5 Release of Data Under the Freedom of Information Act..... 11
7.1.2 Subject Data..... 12
7.1.2.1 Delivery of Requested Subject Data..... 12
7.1.2.2 Rights in Subject Data..... 12
7.1.3 Rights in Non-Subject Data..... 12
7.2 Copyrights..... 13
7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory**..... 13
7.2.2 Copyright License to the Government..... 13
7.2.3 Copyright Statement..... 13
7.3 Trademarks and Service Marks..... 13
7.3.1 Ownership of Trademarks and Service Marks..... 13
7.3.2 Obligation of Employees to Report Trademarks and Service Marks.... 13
7.3.3 Obligation of Collaborators to Notify Each Other..... 13
7.3.4 Responsibility for Filing an Application for Trademark or Service Mark..... 14
7.3.5 License to Use Trademark or Service Mark..... 14
7.4 Subject Inventions..... 14
7.4.1 Obligation to Report Subject Inventions..... 14
7.4.1.1 Collaborators' Instruction to Employees..... 14
7.4.1.2 Timely Invention Disclosure by Inventors..... 14
7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator..... 14
7.4.2 Determination of Subject Inventions..... 14
7.4.3 Title to and Ownership of Subject Inventions..... 14

7.4.4 Filing of Patent Applications..... 15
7.4.4.1 Filing of Patent Applications on Solely Made Inventions..... 15
7.4.4.2 Filing of Patent Applications on Jointly Made Inventions..... 15
7.4.4.3 Preserving Intellectual Property Rights..... 15
7.4.4.4 Filing Deadlines..... 15
7.4.4.5 Copies and Inspection..... 15
7.4.4.5.1 Copies of Prosecution Papers..... 15
7.4.4.5.2 Access to Patent Application File and Right to Make Copies..... 16
7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application..... 16
7.4.5 Nonexclusive License to Subject Inventions..... 16
7.4.5.1 Nonexclusive License Grant..... 16
7.4.5.2 Confirmatory Nonexclusive License Agreement..... 16
7.4.6 Option for Exclusive License to Subject Inventions..... 16
7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement. 17
7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions..... 17
7.4.8.1 Exclusive Licenses and Exclusive License Option..... 17
7.4.8.2 Nonexclusive Licenses..... 17
7.5 Non-Subject Inventions..... 18
7.5.1 Ownership of Non-Subject Inventions..... 18
7.5.2 Rights Under Other Agreements..... 18
7.5.3 No License to Non-Subject Inventions..... 18
7.6 Research License..... 18

Article 8. TANGIBLE PROPERTY..... 18

8.1 Title to Preexisting Tangible Property..... 18
8.2 Tangible Property Purchased by Collaborators to Perform Cooperative Work... 18
8.3 Title to Developed Tangible Property..... 19
8.4 Tangible Property Operational and Disposition Costs..... 19
8.5 Disposal of Tangible Property..... 19

Article 9. LIABILITY..... 19

9.1 Extent of Government Liability..... 19
9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Liability..... 19
9.3 *Force Majeure*..... 20

Article 10. GENERAL PROVISIONS..... 20

10.1 Characteristics of the Agreement..... 20
10.1.1 Entire Agreement..... 20
10.1.2 Severability..... 20
10.1.3 Interpretation of Headings..... 20
10.2 Agreement Between Collaborators..... 20
10.2.1 Governing Laws..... 20
10.2.2 Independent Parties/Entities..... 20
10.2.3 Assignment /Subcontracting..... 20

10.2.4 Disputes..... 21
10.2.4.1Settlement and Resolution..... 21
10.2.4.2Continuation of Cooperative Work..... 21
10.2.5 Waivers..... 21
10.2.6 Use of Name or Endorsements..... 22
10.3 Environment, Safety, and Health..... 22
10.4 U.S. Competitiveness..... 22
10.5 Public Release of This Agreement..... 22

Article 11. MODIFICATIONS AND NOTICES..... 22

11.1 Amendments..... 22
11.2 Termination..... 23
11.2.1 Termination by Mutual Consent..... 23
11.2.2 Unilateral Termination..... 23
11.3 Notices..... 23

Article 12. SURVIVING PROVISIONS..... 24

Article 13. DURATION..... 24

Article 14. SIGNATURES..... 24

Appendix A
Statement of Work

Appendix B
Confirmatory License Agreement

STANDARD
NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

PREAMBLE

Under authority of the U.S. Federal Technology Transfer Act of 1986 (Public Law 99-502, 20 October 1986, as amended), **NAVAIR ORLANDO TSD (Naval Air Warfare Center Training Systems Division)**, located at **12350 Research Parkway, Orlando, Florida 32826-3275**, and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, whose administrative offices are located at **12443 Research Parkway, Suite 207, Orlando, FL 32826-3252**, enter into this Cooperative Research and Development Agreement (CRADA), which shall be binding upon the Collaborators and their assignees according to the clauses and conditions hereof and for the term and duration set forth.

The U.S. Federal Technology Transfer Act of 1986, as amended, provides for making the expertise, capabilities, and technologies of U.S. Federal laboratories accessible to other Federal agencies; units of State or local government; industrial organizations (including corporations, partnerships and limited partnerships, and industrial development organizations); public and private foundations; nonprofit organizations (including universities); or other persons in order to improve the economic, environmental, and social well-being of the United States by stimulating utilization of U.S. Federally funded technology developments and/or capabilities.

NAVAIR ORLANDO TSD has extensive expertise, capabilities, and information in **Virtual Environments for Operator Training**, and, in accordance with the U.S. Federal Technology Transfer Act, desires to make this expertise and technology available for use in the public and private sectors.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory has the interest, resources, capabilities, and technical expertise to transition the results of Naval research and development for public use.

NOW THEREFORE, the Collaborators agree as follows.

Article 1. DEFINITIONS

As used in this Agreement, the following terms shall have the meanings defined below, which are equally applicable to both the singular and plural forms of nouns or any tense of verbs.

1.1 "Agreement" means this Cooperative Research and Development Agreement (CRADA) with its Appendices.

1.2 "Classified Information" means all Data classified in accordance with the national security laws of the United States.

1.3 "Collaborator" means the Navy participant or the Non-Navy participant represented and bound by the signatories of this Agreement.

1.4 "Controlled Unclassified Information (CUI)" means Government Data, Information, or materials provided to or resulting from this Agreement that may be export controlled, sensitive, for official use only, or otherwise protected by law, executive order, or regulation.

1.5 "Cooperative Work" means research, development, engineering, or other tasks performed under this Agreement by **NAVAIR ORLANDO TSD** or **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** working individually or together, pursuant to the Objectives (Article 2) and the Statement of Work (Appendix A).

1.6 "Data" means recorded information of any kind regardless of the form or method of the recording, including computer software.

1.7 "Effective Date" means the date of the last signature of the Collaborators executing this Agreement.

1.8 "Exclusive License" means the grant by the owner of Intellectual Property of the exclusive right to make, use, or sell a patented invention.

1.9 "Government" means the Government of the United States of America.

1.10 "Government Purpose Rights" means the right of the Government to use, duplicate, or disclose Data, in whole or in part, and in any manner, for Government purposes only, and to have or permit others to do so for Government purposes only. Government Purpose Rights includes competitive procurement, but does not include the right to have or permit others to use Data for commercial purposes.

1.11 "Information" means all data, trade secrets, and commercial and financial information. (Chapter 5 Subsection II of Title 5 USC)

1.12 "Intellectual Property" means the property of ideas, examples of which include, but are not limited to, patents, trademarks, copyrights, and trade secrets.

1.13 "Invention" means any invention or discovery that is or may be patentable or otherwise protected under Title 35, United States Code, or any novel variety of plant that is or may be patentable under the Plant Variety Protection Act. (15 USC 3703(9))

1.14 "Invention Disclosure" means the document identifying and describing to organizational management the Making of an Invention.

1.15 "Made" when used in conjunction with any Invention means the conception or first actual reduction to practice of such Invention. (15 USC 3703(10))

1.16 "Militarily Critical Technologies (MCT)" means those technologies identified in the Militarily Critical Technologies List and under the Export Administration Act of 1979, as amended.

1.17 "Non-Subject Data" means any Data that are not Subject Data.

1.18 "Non-Subject Invention" means any Invention that is not a Subject Invention.

1.19 "Patent Application" means an application for patent protection for an Invention with any domestic or foreign patent-issuing authority.

1.20 "Principal Investigator (PI)" means that person having the responsibility for the performance of the Cooperative Work on behalf of a Collaborator.

1.21 "Proprietary Information" means information that embodies trade secrets developed at private expense or business, commercial, or financial information that is privileged or confidential provided that such information:

is not known or available from other sources without obligations concerning its confidentiality;

has not been made available by the owners to others without obligation concerning its confidentiality;

is not already available to the Government without obligation concerning its confidentiality; and

has not been developed independently by persons who have had no access to the information. (FAR/DFARS Definition)

1.22 "Restricted Access Information" means Subject Data generated by **NAVAIR ORLANDO TSD** that would be Proprietary Information if the Information had been obtained from a non-Federal Collaborator participating in a CRADA (15 USC 3710a). Under 15 USC 3710a(c)(7)(B), the Collaborators mutually may agree to provide appropriate protection to Subject Data generated by **NAVAIR ORLANDO TSD** (Restricted Access Information) against public dissemination or release under the Freedom of Information Act (FOIA) for a period of up to five (5) years after development of the Information.

1.23 "Subject Data" means that Data first recorded in the performance of the Cooperative Work.

1.24 "Subject Invention" means any Invention Made in the performance of the Cooperative Work.

1.25 "Tangible Property" means personal or real property that can be physically touched or held.

1.26 "Unlimited Rights" means the right to use, modify, reproduce, release, disclose, perform, or display Data or Computer Programs in whole or in part, in any manner and for any purpose whatsoever, and to have or permit others to do so.

Article 2. OBJECTIVES

Background

NAVAIR ORLANDO TSD has been conducting Navy sponsored research in virtual environments for operator/small arms training for over 15 years. Results of this research have already been transitioned for many military training applications, and have also been the basis for technology transition in use by law enforcement. The Small Arms Simulation Training (SAST) laboratory at **NAVAIR ORLANDO TSD** now has in use an extensive range of software and data bases that have resulted from in-house **NAVAIR ORLANDO TSD** research work.

The University of Central Florida (UCF), with plans to establish a new laboratory for Human Factors research related to Homeland Defense, became aware of the work already done in **NAVAIR ORLANDO TSD's** SAST laboratory. Discussions were immediately begun to explore the mutual benefits of an agreement to work collaboratively.

The primary objective of this work will be to enhance mutual understand of operator performance under stress and develop possible applications of human factors research in this area for military training, and for homeland defense and related technologies

Benefit to Navy Partner: Research using **NAVAIR ORLANDO TSD** developed or enhanced source code may be expected to result in further enhancements to software and databases, at no expense to the government, and with direct use in military simulation-based training. Additionally, the anticipated UCF human factors research projects may be expected to have direct application and benefit for military training. Further, expertise will be provided to UCF on a cost basis, leveraging Navy research dollars.

Benefit to Non-Navy partner: **NAVAIR ORLANDO TSD** developed or enhanced source code for scenario generation, rural and urban databases, weapon tracking, image generation, networking/communication, etc., can provide the kind of virtual environment test bed required for planned human factors research, and can greatly accelerate work toward research objectives.

Article 3. RESPONSIBILITIES

The Collaborators shall provide personnel, facilities, and equipment necessary for, and shall perform, the Cooperative Work.

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities

The Cooperative Work done by **NAVAIR ORLANDO TSD** will be performed under the program guidance of **Ron Wolff, PI, NAVAIR ORLANDO TSD, AIR 4962**, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed

within the facilities of **NAVAIR ORLANDO TSD** or done on behalf of **NAVAIR ORLANDO TSD** by third parties in support of this Agreement.

UCF MURI / OPUS Laboratory personnel who perform Cooperative Work at **NAVAIR ORLANDO TSD** facilities will be supervised by the **UCF MURI / OPUS Laboratory** PI.

3.2 **UCF MURI / OPUS Laboratory** Personnel and Facilities

The Cooperative Work done by the **UCF MURI / OPUS Laboratory** will be performed under the program guidance of **Dr. Peter Hancock, UCF MURI / OPUS Laboratory** PI, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed within the facilities of the **UCF MURI / OPUS Laboratory** or done on behalf of the **UCF MURI / OPUS Laboratory** by third parties in support of this Agreement.

NAVAIR ORLANDO TSD personnel who perform Cooperative Work at the **UCF MURI / OPUS Laboratory** facilities will be supervised by **NAVAIR ORLANDO TSD** PI.

3.3 Security Regulations and Directives

Each Collaborator will abide by the safety and security regulations and directives of the host facility in which the Cooperative Work is being performed.

Article 4. REPRESENTATIONS AND WARRANTIES

4.1 **NAVAIR ORLANDO TSD**'s Representations and Warranties

NAVAIR ORLANDO TSD hereby warrants and represents to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** as follows:

4.1.1 **NAVAIR ORLANDO TSD** is a Federal laboratory of the U.S. Department of the Navy (Navy) as defined by 15 USC 3710a(d)(2)(A) and Department of Defense Instruction 5535.8, dated May 14, 1999.

4.1.2 The performance of the activities specified by this Agreement is consistent with the **Virtual Environments for Operator/Small Arms Training** and technology transfer missions of **NAVAIR ORLANDO TSD** (15 USC 3710a).

4.1.3 The Department of the Navy official executing this Agreement for **NAVAIR ORLANDO TSD** has the requisite power and authority to enter into this Agreement and to bind **NAVAIR ORLANDO TSD** to perform according to the terms of this Agreement.

4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Representations and Warranties

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory hereby warrants and represents to **NAVAIR ORLANDO TSD** as follows:

4.2.1 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not directly or indirectly controlled by a foreign company or government (Executive Order 12591, Section 4 (a)). **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, as of the Effective Date of this Agreement, is a part of a **university** duly organized, validly existing, and in good standing under the laws of **the State of Florida.**

4.2.2 The official executing this Agreement for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** has the requisite power and authority to enter into this Agreement and to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to perform according to the terms of this Agreement.

4.2.3 The University of Central Florida and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** have taken all actions required by law, or otherwise, to authorize the execution and delivery of agreements, such as this Agreement.

4.2.4 The execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any agreement binding on **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. Furthermore, the execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any valid order of any court, or any regulatory agency or other body having authority to which **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is subject.

4.2.5 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not presently subject to debarment or suspension by any agency of the Government. Should **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** be debarred or suspended during the term of this Agreement or thereafter, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** will notify **NAVAIR ORLANDO TSD** within thirty (30) days of receipt of a final notice. **NAVAIR ORLANDO TSD** may then elect to terminate this Agreement and any licenses and options granted under this Agreement.

4.2.6 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is not** a small business as defined in 15 USC 632 and implementing regulations (13 CFR 121.101 et seq.) of the Administrator of the Small Business Administration.

4.3 Joint Representations

The Collaborators make the following representations.

4.3.1 There is no express or implied warranty as to any research, Invention, or product, whether tangible or intangible. In particular, the Collaborators make no express or implied warranty as to the merchantability or fitness for a particular purpose of any research, Invention, or product, whether tangible or intangible. Likewise, the Collaborators make no express or implied warranty as to any Cooperative Work, Subject Invention, Subject Data, or other product resulting from the Cooperative Work.

4.3.2 The use and dissemination of Information and materials exchanged under this Agreement will be in accordance with all U.S. laws and regulations, including those pertaining to national security and export control. Nothing in this Agreement shall be construed as a license to export Information or to permit any disclosure in violation of law, regulation, or Department of Defense policy. The exporting Collaborator is responsible for obtaining any export licenses that may be required by U.S. Federal law.

Article 5. FUNDING

5.1 Payment Schedule

Consistent with the information in Appendix A, **NAVAIR ORLANDO TSD** will provide source code/databases developed or enhanced by **NAVAIR ORLANDO TSD** to the **MURI / OPUS** lab at no cost to **UCF**. If, at a later date during the period of this CRADA, there is a mutually determined need for additional expertise from **NAVAIR ORLANDO TSD**, then the following will apply. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees to pay **NAVAIR ORLANDO TSD** on a cost basis, as mutually determined to be required, during the period of this CRADA, as specified in Appendix A.

Checks will be payable to:

NAVAIR ORLANDO TSD, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments

Each check and its cover correspondence shall refer to Navy CRADA number "NCRADA-**NAVAIR ORLANDO TSD-02-037**."

Checks will be mailed to:

NAVAIR ORLANDO TSD, 12350 Research Parkway (AIR 4.9T Nimmo), Orlando, Florida 32826-3275, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments.

5.2 Insufficient and Excess Funds

NAVAIR ORLANDO TSD may discontinue performance under this Agreement if the funds provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** for performance by **NAVAIR ORLANDO TSD** are insufficient or are not provided as specified in Article 5.1. In the event **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to tender the Government the required payment within fifteen (15) days after its respective due date, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall be in default under this Agreement for failure to make payments. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on**

behalf of the MURI / OPUS Laboratory is in default for this reason, NAVAIR ORLANDO TSD shall notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** does not cure the default within fifteen (15) days of mailing date of notice, **NAVAIR ORLANDO TSD** may proceed to terminate the Agreement in accordance with Article 11.2.2, may cancel any option for an Exclusive License to a Subject Invention, and may terminate any Exclusive License granted pursuant to this Agreement.

Funds that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** paid under Article 5.1 and that **NAVAIR ORLANDO TSD** has not obligated or expended at the time of completion, expiration, or termination of this Agreement shall be returned to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** after **NAVAIR ORLANDO TSD's** submission of a final fiscal report to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**.

5.3 No New Commitments

NAVAIR ORLANDO TSD shall make no new commitments concerning this Agreement after receipt of a written termination notice from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in accordance with Article 11.2 and shall, to the extent practicable, cancel all outstanding commitments by the termination date. Should such cancellation result in any costs incurred by **NAVAIR ORLANDO TSD, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that such costs shall be chargeable against any funding that it provided to **NAVAIR ORLANDO TSD.**

5.4 Accounting Records

NAVAIR ORLANDO TSD shall maintain current accounts, records, and other evidence supporting all its expenditures against funding provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement and shall retain such records for at least twelve (12) months after the completion, expiration, or termination of this Agreement. **NAVAIR ORLANDO TSD** shall provide **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** a financial report within four (4) months after completion, expiration, or termination of this Agreement.

Article 6. REPORTS AND PUBLICATIONS

6.1 Final Report

The Collaborators shall submit to each other a final report within four (4) months of the completion, termination, or expiration of this Agreement that includes the results obtained and a list of all Subject Inventions Made.

6.2 Agreement to Confer Prior to Publication or Public Disclosure

The Collaborators agree to confer and consult prior to any publication or public disclosure of Subject Data to ensure that no Proprietary Information, Restricted Access Information, Government Classified Information, CUI, or MCT Information is released and that patent rights are not compromised. Prior to any such publication or public disclosure of Subject Data, each Collaborator shall be offered a period not less than fifteen (15) days and not to exceed thirty (30) days, unless otherwise mutually agreed in writing by the Collaborators, to review any proposed abstract, publication, presentation, or other document for public disclosure that contains Subject Data. For the purposes of this Article, the term "disclosure" shall include, but not be limited to, submission of any manuscript for peer review prior to publication. It is the responsibility of the Collaborator intending to make public disclosure of Subject Data to notify the other Collaborator of such intent.

If a Collaborator objects to a proposed public disclosure, that Collaborator must so notify the other Collaborator within thirty (30) days of the date of notice of intent to disclose publicly. If no objection is received by the Collaborator intending to make public disclosure, concurrence is assumed. If a Collaborator objects on the grounds that patent rights may be compromised, a Patent Application must be filed by the responsible Collaborator within ninety (90) days of the date of notification of intent to make public disclosure, or by another date mutually agreed to by the Collaborators. If a Collaborator objects to the release of Information on the grounds that the Information is Proprietary Information, Restricted Access Information, or Information whose dissemination is restricted by U.S. security laws or regulations, the disclosure shall be postponed until the Information no longer meets the definitions of Proprietary Information, Restricted Access Information, or is no longer covered by U.S. security laws or regulations.

6.3 Classified Information

Any presentation that includes Subject Data that are Classified Information or otherwise restricted Data must have prior review and approval by **NAVAIR ORLANDO TSD** pursuant to the pertinent security laws, regulations, and directives.

Article 7. INTELLECTUAL PROPERTY

7.1 Data

7.1.1 General Provisions Applying to All Data

7.1.1.1 Ownership

Each Collaborator shall have title to all Data generated by that Collaborator.

7.1.1.2 No Implied License

Unless otherwise specifically provided, the Collaborators agree that the exchange of Data of any kind does not confer a license to any Invention claimed in any

patent or Patent Application or to the subject matter of any copyright, trademark/service mark, or other form of Intellectual Property protection.

7.1.1.3 Marking of Data

7.1.1.3.1 Data Provided With Less Than Unlimited Rights

Each Collaborator shall mark all Data that it provides with less than Unlimited Rights with a marking that clearly identifies the limited rights.

7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall place a proper proprietary marking on each medium used for recording Data that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** delivers to **NAVAIR ORLANDO TSD** under this Agreement that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** asserts is Proprietary Information. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall request in writing if it wishes Subject Data generated by **NAVAIR ORLANDO TSD** to be marked as Restricted Access Information. The Collaborators together shall confer to determine if such marking is appropriate, with reference to the Definitions of Article 1. If the Collaborators mutually agree to the marking then:

(a) For Non-Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory – NAVAIR ORLANDO TSD** MAY USE ONLY FOR PURPOSE OF CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(b) For Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** – GOVERNMENT HAS GOVERNMENT PURPOSE RIGHTS UNDER CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(c) For Data that are Restricted Access Information, the marking shall read:

"RESTRICTED ACCESS INFORMATION – PROTECT IN ACCORDANCE WITH CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**" UNTIL **SEPTEMBER 2005**".

7.1.1.3.3 Data That are Subject to 35 USC 205

NAVAIR ORLANDO TSD shall mark Data it provides under this Agreement that disclose one or more Inventions in which the Government owns or may own a right, title or interest, and that are subject to confidentiality under 35 USC 205. Such Data shall be marked:

"**NAVAIR ORLANDO TSD** DATA PROTECTED FROM RELEASE OR DISCLOSURE UNDER 35 USC 205."

7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted

Each Collaborator shall mark all Data that are Classified Information, CUI, MCT, or otherwise restricted by U.S. security or export control laws or regulations that it provides under this Agreement.

7.1.1.4 Protection of Data

Except for the rights granted in Article 7.1.2.2, Data shall be protected in accordance with the proper markings of its owner and as provided by, at a minimum, the requirements of 15 USC 3710a. Proprietary Information will be protected only if it is properly marked as such. Information provided in intangible form that is Proprietary Information must be designated Proprietary Information at the time it is delivered, followed within fifteen (15) days by a writing summarizing the exact Information to be protected. The Collaborator receiving Information in an intangible form that is designated as Proprietary Information shall be responsible for protecting the Information as Proprietary Information during the fifteen (15) day notification period. After the fifteen (15) day period, if no written summary has been received, the receiving Collaborator need not continue to protect the Information received in intangible form.

Restricted Access Information shall be protected from public dissemination for up to five (5) years, as mutually agreed.

Classified Information, CUI, MCT, or otherwise restricted Information shall be protected in accordance with the security laws of the United States.

7.1.1.5 Release of Data Under the Freedom of Information Act

Data in the possession of **NAVAIR ORLANDO TSD** that are not marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or Restricted Access Information must be released by **NAVAIR ORLANDO TSD** where such release is required pursuant to a request under the Freedom of Information Act (FOIA) (5 USC 552). **NAVAIR ORLANDO TSD** shall protect Data that are properly marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI**

/ **OPUS Laboratory** or Restricted Access Information from release under the FOIA for as long as the marked Data meet the definition of CUI, Proprietary Information or Restricted Access Information. Prior to release of any such Data, **NAVAIR ORLANDO TSD** shall promptly notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** of any request for Data of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** regardless of whether the requested Data are marked Proprietary Information or as required under the public records law.

7.1.2 Subject Data

7.1.2.1 Delivery of Requested Subject Data

Each Collaborator shall have the right to review and receive delivery of all Subject Data generated by the other Collaborator. Requested Subject Data shall be delivered to the requesting Collaborator within fifteen (15) days of the request.

7.1.2.2 Rights in Subject Data

Except as represented in Article 4.3.2, the Collaborators shall have Unlimited Rights in all Subject Data that are not Proprietary Information or Restricted Access Information. Notwithstanding 15 USC 3710a, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** grants Government Purpose Rights in any Subject Data furnished by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to **NAVAIR ORLANDO TSD** under this Agreement that are properly marked as Proprietary Information. The Government has Government Purpose Rights in Subject Data that are Restricted Access Information.

7.1.3 Rights in Non-Subject Data

The Collaborators shall have Unlimited Rights in any Non-Subject Data delivered under this Agreement that are not Proprietary Information.

NAVAIR ORLANDO TSD has a limited right to use, reproduce, and disclose only to Government employees for use in support of the Cooperative Work any Non-Subject Data that are properly marked as Proprietary Information and are provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement. Such Proprietary Information can be used only for the purpose of performing the Cooperative Work unless consent to other use or disclosure is obtained from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in writing.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall have a limited right to use, reproduce, or disclose Non-Subject Data that may describe one or more Inventions in which the Government owns or may own a right, title or interest, if such Non-Subject Data are provided by **NAVAIR ORLANDO TSD** under this Agreement. In accordance with 35 USC 205, such Non-Subject Data are to be held in confidence. Such Non-Subject Data shall be properly marked by **NAVAIR ORLANDO TSD** and the limited rights of **The UCF Office of Research, UCF on behalf of its**

Board of Trustees on behalf of the MURI / OPUS Laboratory shall be defined by a separate non-disclosure agreement.

7.2 Copyrights

7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory may copyright works of authorship prepared pursuant to this Agreement if eligible for copyright protection under Title 17 USC.

7.2.2 Copyright License to the Government

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory grants to the Government a nonexclusive, irrevocable, paid-up license in copyrighted works of authorship, including software (17 USC 106) prepared pursuant to this Agreement for any purpose, consistent with the rights in Data described in Article 7.1.

7.2.3 Copyright Statement

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall include the following statement on any text, drawing, mask work or other work of authorship, that may be copyrighted under 17 USC, that is created in the performance of this Agreement:

"The U.S. Government has a copyright license in this work pursuant to a Cooperative Research and Development Agreement with **NAVAIR ORLANDO TSD.**"

7.3 Trademarks and Service Marks

7.3.1 Ownership of Trademarks and Service Marks

The Collaborator first establishing a trademark or service mark for goods or services with which the mark is used shall be considered the owner of the mark.

7.3.2 Obligation of Employees to Report Trademarks and Service Marks

Employees of both Collaborators shall report the adoption of a trademark or service mark associated with the Cooperative Work to their employer within thirty (30) days of the first use of the mark. Use includes internal use of any product or service of the Cooperative Work.

7.3.3 Obligation of Collaborators to Notify Each Other

Each Collaborator shall notify the other Collaborator within thirty (30) days of their employee's report of the first use of a trademark or service mark.

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

Each Collaborator shall be entitled to own the Subject Inventions of its employees. Each Collaborator shall cooperate with the other Collaborator to obtain inventor signatures on Patent Applications, assignments or other documents required to secure Intellectual Property protection. For any Invention Made jointly by employees of the Collaborators, each Collaborator shall have ownership of the Subject Invention in the form of an undivided interest.

7.4.4 Filing of Patent Applications

7.4.4.1 Filing of Patent Applications on Solely Made Inventions

Each Collaborator has primary responsibility for filing Patent Applications on the Subject Inventions of its employee(s).

Notwithstanding such primary responsibility, by mutual agreement, the Collaborators may identify which Collaborator shall file a Patent Application on any Subject Invention.

7.4.4.2 Filing of Patent Applications on Jointly Made Inventions

In the case of an Invention jointly Made by employees of both Collaborators, the Collaborators shall confer and agree as to which Collaborator will file any Patent Application. Officers of the non-filing Collaborator shall cooperate with the filing Collaborator to obtain signatures on documents that are needed to file a Patent Application.

7.4.4.3 Preserving Intellectual Property Rights

The Collaborator responsible for filing of a Patent Application on any Subject Invention shall file such Patent Application at least sixty (60) days prior to any bar date or one year from the date the Invention Disclosure was received, whichever comes first. If no Patent Application is filed within the specified time period, the other Collaborator may assume control of filing the Patent Application and take title to the Subject Invention on ten (10) days written notification. The Collaborator that relinquished the responsibility to file shall retain a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Subject Invention practiced throughout the world by or on its behalf.

7.4.4.4. Filing Deadlines

The Collaborator responsible for filing any Patent Application for a Subject Invention shall notify the other Collaborator of all filing deadlines for prosecution of any Patent Application and maintenance of any patents on the Subject Invention. Notwithstanding the primary responsibility defined in Article 7.4.4.1, sixty (60) days prior to any filing deadline, the Collaborators shall confer to determine if the filing Collaborator intends to respond to the filing deadline. The non-filing Collaborator will be permitted to take action if the filing Collaborator declines.

7.4.4.5 Copies and Inspection

7.4.4.5.1 Copies of Prosecution Papers

Each Collaborator filing a Patent Application on a Subject Invention shall provide the other Collaborator with a copy of any communication relating

to prosecution of said Patent Application within thirty (30) days of receipt of such communication.

7.4.4.5.2 Access to Patent Application File and Right to Make Copies

Upon written request, the filing Collaborator shall give the other Collaborator an Associate Power of Attorney, with authorization to access the Patent Application, make copies, and, in the event the filing Collaborator fails or declines to take action, do all that is necessary to secure Intellectual Property protection for the Subject Invention.

7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application

In the event both Collaborators decline to file a Patent Application on a Subject Invention, the Government will renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

In the event both Collaborators decline to file a Patent Application on a Subject Invention, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** may, at its sole discretion, renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

7.4.5 Nonexclusive License to Subject Inventions

7.4.5.1 Nonexclusive License Grant

Each Collaborator grants to the other Collaborator a nonexclusive, irrevocable, paid-up license to practice a Subject Invention Made by employees of the granting Collaborator or have the Subject Invention practiced throughout the world by or on behalf of the other Collaborator. No nonexclusive license granted under this Agreement shall permit licensee to grant sublicenses.

7.4.5.2 Confirmatory Nonexclusive License Agreement

Each Collaborator has the obligation to provide a Confirmatory License Agreement (Appendix B) to the other Collaborator for each nonexclusive license within ninety (90) days of the date of filing.

7.4.6 Option for Exclusive License to Subject Inventions

NAVAIR ORLANDO TSD gives **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** the option of acquiring an Exclusive License for the field of use **(virtual environments for operator/small arms training)** in the Government's rights in any Subject Invention Made in whole or in part by a **NAVAIR ORLANDO TSD** employee. The license shall be for reasonable consideration. In

order to exercise this option, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must notify **NAVAIR ORLANDO TSD** in writing within one hundred and eighty (180) days of the filing of a Patent Application. Unless another time period is mutually agreed upon between the Collaborators, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must execute an Exclusive License to the Subject Invention within one hundred and eighty (180) days of election to exercise the option, or the Invention shall be made available for licensing by the public in accordance with 37 CFR Part 404.

Any Exclusive License granted by the Government in a Subject Invention is subject to the statutorily required reservation by the Government of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the Government (15 USC 3710a).

7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement

No license granted under this Agreement shall be assigned, licensed or otherwise disposed of except to the successor in interest of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which such license pertains.

7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions

7.4.8.1 Exclusive Licenses and Exclusive License Option

NAVAIR ORLANDO TSD may terminate any Exclusive License or cancel any option for an Exclusive License to a Subject Invention granted under this Agreement in the event that:

(a) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is in default for failure to make payment as agreed in Article 5; or

(b) The Agreement is terminated unilaterally by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**; or

(c) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to perform according to the Statement of Work (Appendix A); or

(d) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** becomes a foreign owned, controlled, or influenced (FOCI) organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.4.8.2 Nonexclusive Licenses

NAVAIR ORLANDO TSD shall terminate any nonexclusive license to a Subject Invention granted under this Agreement if **The UCF Office of Research,**

UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory becomes a FOCI organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.5 Non-Subject Inventions

7.5.1 Ownership of Non-Subject Inventions

Each Collaborator owns its Non-Subject Inventions.

7.5.2 Rights Under Other Agreements

Nothing in this Agreement is intended to change the rights in Intellectual Property acquired by the Collaborators in any other contract or agreement between the **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** and the Government.

7.5.3 No License to Non-Subject Inventions

This Agreement does not grant any Collaborator a license, express or implied, to any Non-Subject Invention.

7.6 Research License

Each Collaborator shall allow the other Collaborator to practice any of its Non-Subject Inventions for the purpose of performing the Cooperative Work.

No license, express or implied, for commercial application(s) is granted to either Collaborator in Non-Subject Inventions by performing the Cooperative Work.

For commercial application(s) of Non-Subject Inventions, a license must be obtained from the owner.

Article 8. TANGIBLE PROPERTY

8.1 Title to Preexisting Tangible Property

Each Collaborator shall retain title to all Tangible Property to which it had title prior to the Effective Date of this Agreement.

8.2 Tangible Property Purchased by Collaborators to Perform the Cooperative Work

Each Collaborator shall retain title to all Tangible Property that it purchases during the period of this Agreement. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** cannot take title to any Government Tangible Property under this Agreement. Collaborator consumables to be used in the Cooperative Work of this Agreement are the property of the purchasing Collaborator until consumed.

8.3 Title to Developed Tangible Property

All Tangible Property developed under this Agreement with all components purchased by one Collaborator shall be the property of that Collaborator. Tangible Property having any component purchased by **NAVAIR ORLANDO TSD** shall be the property of the Government, unless such Tangible Property can reasonably be separated without damage to the other individual components. After this Agreement is completed, expired, or terminated, if separation of components can be made without damage, the Collaborators may, by mutual agreement, separate the Tangible Property into its components and the separated components shall remain the property of the Collaborator that purchased them.

8.4 Tangible Property Operational and Disposition Costs

During the period of and upon completion, expiration, or termination of this Agreement, each Collaborator shall be responsible for all costs of maintenance, removal, storage, repair, disposal, and shipping of all Tangible Property to which it has title.

8.5 Disposal of Tangible Property

Disposal of Tangible Property shall be in accordance with applicable U.S. Federal, State, and local property disposal laws, environmental laws, and regulations.

Article 9. LIABILITY

9.1 Extent of Government Liability

The Government shall be solely liable for the negligent or wrongful acts of its officers and employees to the extent provided for in the Federal Tort Claims Act (28 USC 2671 et. seq.) and in other applicable laws and regulations of the United States that specifically waive sovereign immunity. Nothing in this Agreement shall be construed as a waiver of the sovereign immunity of the United States.

9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Liability

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is solely responsible for its actions and the actions of those acting for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in the performance of this Agreement and for any damages that may arise from any suit, action, or claim, and for any costs from or incidental to any suit, action, or claim, including but not limited to settlement and defense costs. Further, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that in any suit, action or claim brought by anyone not a party to this Agreement based on actions of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not pursue any actions to enter the Government as a party in such suit, action or claim unless the Government has some liability under the Federal Tort Claims Act.

9.3 *Force Majeure*

No Collaborator shall be liable for the consequences of any *force majeure* that (1) is beyond its reasonable control; (2) is not caused by the fault or negligence of such Collaborator; (3) causes such Collaborator to be unable to perform its obligations under this Agreement; and (4) cannot be overcome by the exercise of due diligence. In the event of the occurrence of a *force majeure*, the Collaborator unable to perform shall promptly notify the other Collaborator. The Collaborators shall suspend performance only for such period of time as is necessary to overcome the result(s) of the *force majeure* and shall use their best efforts to resume performance as quickly as possible.

Article 10. GENERAL PROVISIONS

10.1 Characteristics of the Agreement

10.1.1 Entire Agreement

This Agreement constitutes the entire agreement between the Collaborators concerning the Cooperative Work and supersedes any prior understanding or written or oral agreement relative to the Cooperative Work.

10.1.2 Severability

The illegality or invalidity of any Article of this Agreement shall not impair, affect, or invalidate any other Article of this Agreement.

10.1.3 Interpretation of Headings

Headings of the Articles of this Agreement are for convenience of reference only and do not form a part of this Agreement and shall in no way affect the interpretation thereof.

10.2 Agreements Between Collaborators

10.2.1 Governing Laws

United States Federal Laws shall govern this Agreement for all purposes.

10.2.2 Independent Parties/Entities

The relationship of the Collaborators to this Agreement is that of independent parties and not as agents of each other, partners, or participants in a joint venture. Each Collaborator shall maintain sole and exclusive control over its personnel and operations.

10.2.3 Assignment/Subcontracting

10.2.3.1 Neither Collaborator may allow third parties to perform any part of the Cooperative Work under this Agreement without express written consent of the other

Collaborator. If consent is obtained, the Collaborator requesting such consent shall remain fully responsible for the portion of the Cooperative Work to be accomplished under a third-party agreement, and the third party is not a Collaborator of this Agreement. Any third-party agreement to perform a portion of the Cooperative Work shall contain terms consistent with this Agreement.

10.2.3.2 This Agreement shall not be assigned or otherwise transferred by either Collaborator without the prior written consent of the other Collaborator, except to the successor of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which this Agreement pertains.

10.2.3.3 If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee is a U.S. company, and becomes, during the term of this Agreement or thereafter, directly or indirectly owned, controlled, or influenced by a foreign company or government (FOCI), then **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee shall promptly notify **NAVAIR ORLANDO TSD** to that effect.

10.2.4 Disputes

10.2.4.1 Settlement and Resolution

NAVAIR ORLANDO TSD and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agree to use reasonable efforts to reach a fair settlement of any dispute. If such efforts are unsuccessful, remaining issues in dispute will be referred to the signatories or their successors for resolution. If a dispute continues, the remaining issues may be submitted to the Chief of Naval Research (CNR), or the CNR designee, for resolution. This Agreement does not prevent any Collaborator from pursuing disputes in a U.S. Federal court of competent jurisdiction. No Collaborator will pursue litigation in a U.S. Federal court until after the CNR, or the CNR designee, decides the dispute, or until sixty (60) days after the dispute was first submitted to the CNR, or the CNR designee, whichever comes first.

10.2.4.2 Continuation of Cooperative Work

If payments or installment payments are to be made as stated under Article 5, **NAVAIR ORLANDO TSD** will not start or continue cooperative work until payments or installment payments are received.

10.2.5 Waivers

None of the provisions of this Agreement shall be considered waived by either Collaborator unless such waiver is given in writing to the other Collaborator, signed by the executing official of this Agreement or the official's successor having the authority to bind the Collaborator making the waiver. The failure of either Collaborator to insist upon strict performance of any of the terms and conditions herein, or failure or delay to exercise any rights provided herein or by law shall not be deemed a waiver of any right of either Collaborator under this Agreement.

10.2.6 Use of Name or Endorsements

Except as provided for in Article 7.2.3, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not use the name of **NAVAIR ORLANDO TSD** or any other Government entity on any product or service that is directly or indirectly related to either this Agreement or any patent license or assignment associated with this Agreement without the prior approval of **NAVAIR ORLANDO TSD**. By entering into this Agreement, **NAVAIR ORLANDO TSD** does not directly or indirectly endorse any product or service provided, or to be provided, by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, its successors, assignees, or licensees. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not in any way imply that the Department of the Navy endorses any such product or service.

10.3 Environment, Safety, and Health

Each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste in its custody during the course of this Agreement. At the conclusion of this Agreement, each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste still in its possession. Each Collaborator shall obtain at its own expense all necessary permits and licenses as required by U.S. Federal, State, and local law and shall conduct such handling, control, and disposition in a lawful and environmentally responsible manner. Each Collaborator is responsible for all required environmental, safety, and health compliance, notice, and monitoring related to its facility in accordance with U.S. Federal, State, and local law and regulations. Collaborators shall abide by the environmental, safety, and health directives of the host facility in which the Cooperative Work is being performed, and any U.S. Federal, State, or local laws and regulations pertaining to environment, safety, and health that are applicable to the host facility.

10.4 U.S. Competitiveness

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory agrees that any product, process, or service using Intellectual Property arising from the performance of this Agreement shall be manufactured substantially in the United States.

10.5 Public Release of This Agreement

This Agreement, without funding information (Article 5) and Appendices, may be released to the public.

Article 11. MODIFICATIONS AND NOTICES

11.1 Amendments

If a Collaborator wishes to modify this Agreement, the Collaborators shall confer in good faith to determine the desirability of such modification. Such modification shall not be effective until a written amendment is signed by both executing officials of this Agreement or their successors.

11.2 Termination

11.2.1 Termination by Mutual Consent

The Collaborators may elect to terminate this Agreement at any time by mutual consent. Such termination shall not be effective until a written termination agreement is signed by both executing officials of this Agreement or their successors.

11.2.2 Unilateral Termination

A Collaborator may unilaterally terminate this entire Agreement at any time by giving the other Collaborator written notice signed by the executing official of this Agreement or his/her successor, not less than thirty (30) days prior to the desired termination date. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** unilaterally terminates this Agreement, any option for an Exclusive License to a Subject Invention and any Exclusive License to a Subject Invention granted by or pursuant to this Agreement shall simultaneously be terminated.

11.3 Notices

All notices pertaining to or required by Articles of this Agreement, except those pertaining solely to the prosecution of any patent, trademark, or service mark, shall be in writing and shall be signed by an authorized representative of the Technology Transfer Office for **NAVAIR ORLANDO TSD** or the preferred contact for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, and all such notices shall be delivered by hand, sent by courier with proper registration, or sent by certified mail, return receipt requested, with postage prepaid, addressed as follows:

If to **NAVAIR ORLANDO TSD**:

Ms. Patricia R. Nimmo
Technology Transfer Project Manager
NAVAIR ORLANDO TSD (AIR 49T)
12350 Research Parkway
Orlando, FL 32826-3275

If to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

Ms. Kim Smith
Senior Contract Specialist
Office of Research
University of Central Florida
Orlando Tech Center
12443 Research Parkway, Suite 207
Orlando, Fl 32826-3252

A Collaborator shall notify the other Collaborator of a change of address in the manner set forth above.

Notices pertaining solely to the prosecution of any patent, trademark, or service mark related to this Agreement shall be in writing and shall be signed by and sent to the Collaborator's legal counsel for Intellectual Property. Legal counsel for Intellectual Property for each Collaborator shall send a copy of any such notice to the Technology Transfer Office for he **NAVAIR ORLANDO TSD**. If either Collaborator fails to identify such counsel upon request, then such notices shall be sent to the points of contact specified above.

Article 12. SURVIVING PROVISIONS

The Articles covering Definitions, Representations and Warranties, Funding, Reports and Publications, Intellectual Property, Tangible Property, Liability, General Provisions, Modifications and Notices, and Surviving Provisions shall survive the completion, termination, or expiration of this Agreement.

Article 13. DURATION

This Agreement expires three years after its Effective Date, unless otherwise extended in writing according to the provisions of Article 11.

Article 14. SIGNATURES

For **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

I, the undersigned, am duly authorized to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002

By: M.J. Soileau

Title: Vice President, Office of Research

For the Department of the Navy:

I, the undersigned, by 15 USC 3710a and Navy regulations, am duly authorized to bind the U.S. Navy to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002.

By:

R. A. MOHLER

Title: Captain, U.S. Navy
Commanding Officer

Navy Organization: NAVAIR ORLANDO, TRAINING SYSTEMS DIVISION

This Page Intentionally Left Blank

APPENDIX A

STATEMENT OF WORK
BETWEEN
NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND
THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

NAVAIR ORLANDO TSD will:

1. Provide, for use in the UCF MURI/OPUS laboratory, source code/databases developed or enhanced by NAVAIR ORLANDO TSD as part of Navy sponsored research and development, e.g., Host Software, Image Generator Software, Weapon Tracking Software, Networking/Communications Software, Data Collection/Analysis Software, Scenario Builder Software, Rural and Urban Databases

2. Provide expertise, on a cost basis and as mutually determined to be required, on utilization of government provided software and databases for UCF MURI/OPUS Laboratory research

3. Provide, on a cost basis and as mutually determined to be required, expertise for further hardware/software/database development, testing, and training applications

4. Participate, on a cost basis, in mutually selected specific research projects conducted under the auspices of the UCF MURI/OPUS Laboratory

5. Provide access to NAVAIR ORLANDO TSD customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

3. Mutually develop processes separate from this agreement to allow exchange of funds, should that be mutually determined to be required during the period of this CRADA

NOTE 1: The parties agree that Article 9.2 Extent of Liability, will be subject to and interpreted in accordance with Florida Statutes, Section 768.28.

NOTE 2: FOREIGN ACCESS TO TECHNOLOGY

All work described in this SOW shall be subject to the following with regard to foreign access to technology.

The Collaborators agree that research findings and technology developments under this CRADA may constitute a significant enhancement to the national defense. Accordingly, access to important technology developments under this CRADA by foreign firms or institutions must be carefully controlled. These controls are in addition to, and are not intended to change or supercede, the provisions of the International Traffic in Arms Regulation (22 CFR pt. 121 et seq.) the DoD Industrial Security Regulation (DoD 5220.22-R) and the Department of Commerce Export Regulation (15 CFR pt.770 et seq.)

The non-Navy Collaborator shall provide timely notice to the government's designated Principal Investigator of any proposed transfer of research findings and technology developments under this CRADA to foreign firms or institutions. If the government determines that the transfer may have adverse consequences to the national security interests of the United States, the non-Navy Collaborator and the government shall jointly endeavor to find alternatives to the proposed transfer which obviate or mitigate potential adverse consequences of the transfer but which provide substantially equivalent benefits to the non-Navy Collaborator.

APPENDIX B

CONFIRMATORY LICENSE AGREEMENT	*1. APPLICATION FOR (Title of Invention)*
2. INVENTOR(S) AND AFFILIATION	
3. PATENT APPLICATION SERIAL NO.	*4. PATENT APPLICATION FILING DATE*
5. NAVY ACTIVITY (Name, address, point of contact)	*6. NON-NAVY ACTIVITY (Name, address, point of contact)*
7. CRADA AGREEMENT NO.	*8. DATE OF THIS AGREEMENT*

9. The Invention identified above is a "Subject Invention" under Article 7 Intellectual Property included with the CRADA identified in Box 7 between the Department of the Navy and Non-Navy Activity identified in Box 6.

This document is confirmatory of the nonexclusive, irrevocable, paid-up license to practice the identified Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the receiving party, and of all other rights acquired by the receiving party by the referenced clause.

This license is granted to

_____ *the Government*

(Select one)

_____ *Non-Navy Activity identified in Box 6*

under this CRADA in the identified Invention, Patent Application and any resulting patent.

The licensee is hereby granted an irrevocable power to inspect and make copies of the above-identified Patent Application.

_____ *ACTIVITY NAME OF LICENSOR*

_____ *SIGNATURE*

_____ *NAME (Typed or Printed)*

_____ *TITLE*

_____ *BUSINESS TELEPHONE*

University of Central Florida

Multi-Disciplinary University Research Initiative

MURI-OPUS

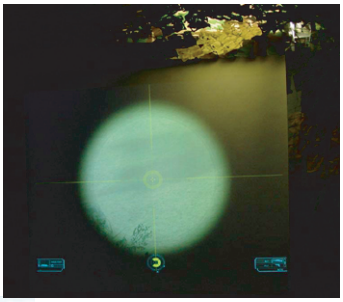
Operator Performance Under Stress

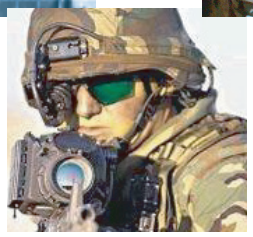
www.mit.ucf.edu

VOLUME THREE











2003 Executive Summary

Table of Contents

NCRADA

A Collaborative Partnership between:

University of Central Florida

Contact Information:

University of Central Florida • MIT^2 Laboratory

3100 Technology Pkwy, Suite 337 • Partnership II • Orlando, Florida 32826-0544

http://www.mit.ucf.edu

Executive Summary

Accomplishments to Date

We are approaching the end of the third year of the Multiple University Research Initiative (*MURI*) program on Operator Performance Under Stress (*OPUS*). This report represents an account of progress to date, which contains numerous successes achieved at this stage. Following the original proposal, we have elaborated upon a number of areas that remain unresolved about the action of stress on performance. Each of these issues has served as a central theme in our on-going series of 'White Papers' that act to guide empirical research efforts. The three completed 'White Papers' are included in the present report and are to be followed by two additional 'White Papers' that, added to our initial model, will form the five pillars of our theoretical foundation. One of these white papers is near completion, and the last white paper is still in its early stages. Although we have detailed the progress of each discrete section of the MURI program individually, we continue to strengthen and exploit each of the theoretical, methodological, and electronic links between our respective laboratories. At the University of Central Florida, the MIT^2 research group has made considerable progress. We have been able to identify, generate, acquire and use our research test beds to evaluate our respective theoretical propositions. From the work done so far we now know much more about variations of spatio-temporal perception under stressful conditions. In addition, initial experiments testing Fuzzy Signal Detection Theory (FSDT) indicate that this novel approach to performance measurement captures elements of performance in complex tasks with high uncertainty and is a useful extension of traditional Signal Detection Theory. Our laboratory has also examined the role of individual differences in adaptation to stress associated with perceptual and cognitive tasks. Evidence to date indicates that personality traits of individuals influence their stress level and the coping strategies they employ to deal with the stress. We have recently completed a 'proof of concept' meta-analytic study of the effect of noise on performance on memory tasks. This serves as a foundation for future meta-analytic work on the stress-performance relation. These efforts are geared toward enhancing IMPRINT capacity for prediction of performance under stress. In addition, we have recently extended our field research to shooting performance at a police training shooting range. This facility allows us to measure stress in realistic environments similar to those of military soldiers in urban environments. Through a Cooperative Research and Development Agreement (*CRADA*) with the Naval Air Warfare Center, Training systems Division, Orlando, the MIT^2 laboratory has completed installation of the Small Arms Simulation Test Bed (SAST) facility. We are currently using this facility for empirical investigation of cognitive changes that occur in a stressful target acquisition task. In addition, we are currently working with ARL HRED Laboratory on a possible collaborative simulation research which will upgrade the SAST and allow members of distributed teams to simultaneously participate in a common scenario. When linked together the systems of these two institutions can be used to examine the performance of larger teams. This initiative will allow the expansion of stress research to include social components that follow team performance under stressful situations.

Following our form of interaction, the chief member of the University of Minnesota group is now working at the University of Central Florida. The *VIKING* experimental environment developed at the University of Minnesota has been used to assess performance on tasks that impose information processing demands similar to those faced by soldiers in combat environments. Further development work on the *VIKING* platform is underway at the MIT^2 laboratory at the University of Central Florida. The recently developed ARES battery has been

used to assess the changes in cognitive performance of soldiers in the field during simulated combat, finding that many soldiers tolerate such field conditions with minimal impairment of cognitive performance if they know what is expected but that high demand events are still associated with cognitive performance changes. In addition, it appears that different forms of cognitive task are not affected equally in the context of combat. In addition, cases were observed in that study that clearly showed a precipitous decline in adaptive function at the extremes of stress, as predicted by the Hancock and Warm (1989) model.

The Team Performance Laboratory (*TPL*) at the University of Central Florida has investigated adaptability of teams under time pressure and high workload, and is currently using a military-based task (*Flashpoint*) to further investigate the impact of stress on team processes. In particular, initial evidence indicates that the quality of team communication and performance varies as a function of team type (e.g., command and control versus a reconnaissance team). TPL researchers have also found that time pressure impairs communication among team members (breaking the closed-loop communication pattern). MURI researchers at Kansas State University have investigated effects of trust on team decision-making. These efforts have resulted in the identification of a viable research paradigm for the investigation of trust and team problem solving and represents one of the myriad examples of cross Laboratory technology transfer and protocol integration. Further, the KSU group has conducted three simulated combat environment experiments on the effects of social interaction on performance. Specifically, they have investigated team problem solving and cooperative performance in distributed teams. MURI researchers in the Cognitive Science Laboratory at the Catholic University of America have completed their initial studies on the impact of automation on performance in decision-making tasks of the kind typically faced by military personnel in command and control settings. They found that as a result of complacency, automation with high (but not perfect) reliability does not always support performance. They also found that automation unreliability has a greater performance cost for decision automation than for information automation. In addition, they are exploring the effectiveness of adaptive automation using real-time measurement of workload under conditions of sustained workload and perceived stress. They are currently working on examining blood flow as an index of cognitive workload using the Transcranial Doppler Sonography System (TCDS). System sensitivity to blood flow changes in the left and right middle cerebral arteries is being examined in a working memory task involving storage and recall of digits and in the tracking sub-task of the Multi-Attribute Task Battery. Researchers at the University of Cincinnati have conducted experiments demonstrating that the decline in sustained attention performance over time is accompanied by a parallel decline in cerebral bloodflow in the right hemisphere. In addition, they have initial evidence that transcranial oximetry (TCCO) measures of blood flow and oxygenation levels in the brain may also provide a useful physiological measure of sustained attention.

In respect of all these achievements, MURI has supported over thirty researchers, including three post-doctoral researchers and over twenty graduate student researchers across all participating Institutions.

Contributions to Scientific Understanding

We have made several fundamental contributions to the scientific understanding of stress effects. We have begun to elucidate the effects of diverse sources of real-world stress on spatio-temporal perception. Using stress sources that will impact further development of the IMPRINT model, we have begun to confirm our hypotheses on the 'narrowing' of the perceptual world of

the stressed operator. These efforts have burgeoned into display design guidelines for such operations as a deliverable product, which has been published in *Ergonomics in Design*. We have made substantive theoretical progress on the question of adaptive interfaces for stressful operations. In particular, we have formulated a number critical questions about the efficacy of design and the possibility of optimal design. These developments have helped characterize and strengthen currently tested empirical propositions that promise to save significant dollars against a vision of endless, iterative search for design improvement. Our progress has been facilitated by naturalistic field observations and evaluations. The opportunity to conduct real-world test and evaluation on stressed soldiers and police officers accords with our theoretical proposition that a fully articulated theory can only emerge from an understanding of actual operations. While we continue to pursue laboratory and simulation efforts, we look to a union of forms of observation against which to test hypotheses formulated in the respective 'White Papers.' Our experiments on combined physical and cognitive effort have resulted in a surprising and important discovery. We call this the 'table-top' tolerance limit phenomenon. This phenomenon indicates that even when stress levels are significantly high the performance remains relatively constant to a point of sudden incipient failure. This finding has significant implications for both our future work and planning for military operations in general. Identifying symptomatic characteristics at the edge of failure (e.g., progressive increase in performance variability) may provide a predictive marker that enables the identification of failing operators. This would prove to be a crucial theoretical and practical discovery. We are also now beginning to distill ways in which team performance under stress is not simply the product of individual contributions. In particular, we are starting to show how workload sharing and task division and reallocation by team units acts to diffuse unacceptably high levels of demand. We are using this insight as a basis for designs in adaptive interfaces for stress in which demand may be shared between exposed soldiers and support personnel in remote locations. Initial developments in our collective work demonstrate how direct monitoring of operator state may be mapped into task resolution, well beyond the traditional conception of 'adaptive' interfaces. These new, neuro-ergonomic insights present the next generation level of conceptions for human-machine interaction. We have also found that stress influences mutual trust levels in combined decision-making conditions. The continuation of this effort seeks to explicate the specific mechanism of stress-induced deterioration. These represent only a selection of insights derived from the contained work, others of which are articulated in the specific papers.

Research for the Coming Year

In the coming year we look to complete the 'White Papers' as our theoretical efforts look toward reaching fruition. One of these papers is near completion, and the fifth White Paper will serve as a summary/integration of the previous four papers. Programmatic research evaluations related to the propositions set forth in the white papers will be pursued. Further experimentation is planned on the effects of time-pressure, cognitive workload change, task characteristics, and the degree of operator's control over the task environment on a variety of performance capacities. This experimentation permits tests of our theoretical models of the spatial and temporal components of stress and task performance, as well as the recruitment, allocation, depletion of resources and recovery under conditions of high workload and stress. Further, we plan further evaluation of target acquisition capacities and decision strategies as tests of fuzzy signal detection under stress. For this effort we have begun experimenting with basic shapes but also with more sophisticated and complex realistic objects and events. These studies will complement the psychophysical experiments already in progress and provide further evidence for the practical

Table of Contents

MURI Divisions

University of Central Florida (Peter Hancock)

Supported Work

The Future of Neuroergonomics - *Refereed Publication*
Hancock, P.A. & Szalma, J.L. (2003). The future of Neuroergonomics. *Theoretical Issues in Ergonomic Science*, **44** (1-2), 238-249.

Operator Performance under Stress – *Presentation* ARO Workshop Cashiers *Hancock, P. A. (2003)*

Adaptive Experts and Automaticity: Explanations of Failing Under Pressure *Presentation* ARO Workshop Cashiers *Ward, P., Szalma, J.L., & Hancock, P.A (2003)*

FSDT Initial Evidence for a Novel Approach to Evaluation of Decision-Making Under Uncertainty *Abstract* ARO Workshop Cashiers *Szalma, J.L., &., Hancock, P.A. (2003)*

Stress, Human Information Processing, and Performance Mediators: Application to a Descriptive Framework to Current Modeling Tools *White paper Ward, P., Szalma, J.L., and Hancock, P.A. (2003)*

Performance, Workload, and Stress with Temporal and Spatial Task Demands *Poster* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society: Contemporary and Emerging Issues in Human Factors, Engineering and Military Psychology *Ross, J.M., Szalma, J.L., Thropp, J. & Hancock, P.A. (2003)*

Performance, Workload, and Stress with Temporal and Spatial Task Demands *Paper* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society: Contemporary and Emerging Issues in Human Factors, Engineering and Military Psychology *Ross, J.M., Szalma, J.L., Thropp, J. & Hancock, P.A. (2003)*

Comparison of Methods of Analyses from Fuzzy & Traditional Signal Detection Theory *Poster* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society: Contemporary and Emerging Issues in Human Factors, Engineering and Military Psychology *Murphy, L.L., Szalma, J.L., Hancock, P.A., & Stafford, S.C. (2003)*

Operator Performance Under Stress: Current Progress and Future Prospects *Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society: Contemporary and Emerging Issues in Human Factors, Engineering and Military Psychology *Hancock, P.A. (2003*)

Predicting Performance Under Stressful Conditions: A Theoretical and Empirical Approach *Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society: Contemporary and Emerging Issues in Human Factors, Engineering and Military Psychology *Ward, P., Szalma, J.L., & Hancock, P.A. (2003)*

Application of Fuzzy Signal Detection Theory to the Evaluation of Performance Under Stress *Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society: Contemporary and Emerging Issues in Human Factors, Engineering and Military Psychology *Szalma, J.L., Ward, P., & Hancock, P.A. (2003)*

The Role of Stress and Individual Differences in Warning Compliance *Poster* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society: Contemporary and Emerging Issues in Human Factors, Engineering and Military Psychology *Helmick, J.A, Burke, K.A. & Weaver, J.L. (2003)*

The Application of a Commercially Available Software Program as a Testbed for Stress and Performance *Demonstration at* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society: Contemporary and Emerging Issues in Human Factors, Engineering and Military Psychology *Burke, K.A., Greenwood-Ericksen, A., & Weaver, J.L. (2003)*

Operator Stress and Display Design - *Refereed Publication*
Hancock, P.A., & Szalma, J.L. (2003). Operator stress and display design. *Ergonomics in Design*, **11** (2), 13-18.

Effects of Automation Reliability on Human Monitoring Performance *Proceedings* 47th Annual Human Factors and Ergonomics *Oakley, B., Mouloua, M., & Hancock, P.A. (2003)*

The Role of Choice and Individual Differences in the Mitigation of Noise and Task Stressors on Vigilance *Proceedings* 47th Annual Human Factors and Ergonomics *Ganey, H.C.N., Szalma, J.L., Hancock, P.A., Mouloua, M., Connelly, A., Dalton, J., & Davis, F. (2003)*

Individual Differences in Dispositional Pessimism, Stress, and Coping as a Function of Task Type *Proceedings* 47th Annual Human Factors and Ergonomics *Thropp, J.E., Szalma, J.L., Ross, J.M., & Hancock, P.A. (2003)*

An Affordable Medium-fidelity Virtual Environment Soldier Simulator Designed for Infantry Combat Research *Proceedings* 47th Annual Human Factors and Ergonomics *Greenwood-Ericksen, A., Weaver, J.L., Burke, K., Bradley, K.A., & Hancock, P.A. (2003)*

Comparison of Fuzzy Signal Detection and Traditional Signal Detection Theory: Approaches to Performance Measurement *Proceedings* 47th Annual Human Factors and Ergonomics *Murphy, L., Szalma, J.L., & Hancock, P.A. (2003)*

Comparison of Fuzzy Signal Detection and Traditional Signal Detection Theory: Approaches to Performance Measurement *Poster* 47th Annual Human Factors and Ergonomics *Murphy, L., Szalma, J.L., & Hancock, P.A. (2003)*

Performance, Workload, and Stress with Temporal and Spatial Task Demands *Proceedings* 47th Annual Human Factors and Ergonomics Society *Ross, J.M., Szalma, J.L., Thropp, J., & Hancock, P.A. (2003)*

The Stress Performance Relationship and the Effect of Performance Moderators *Presentation at the* 47th Annual Human Factors and Ergonomics Society *Ward, P., Szalma, J.L. & Hancock, P.A. (2003)*

The Stress-Performance Relationship and the Effect of Performance Mediators: Evidence from a Meta-Analysis of Noise and Memory *Poster at the* 47th Annual Human Factors and Ergonomics Society *Ward, P., Szalma, J.L., Hancock, P.A., Terrence, P.I., Panganiban, A.R., & Morgan, J.F. (2003)*

Applications of Fuzzy Signal Detection Theory to Vigilance: The Effect of Criterion Shifts *Proceedings* 47th Annual Human Factors and Ergonomics Society *Stafford, S.C., Szalma, J.L., Hancock, P.A. & Mouloua, M. (2003)*

An Experimental Investigation of the Skill, Rule, and Knowledge-base Performance under Noise Conditions *Proceedings* 47th Annual Human Factors and Ergonomics Society *Weaver, J.L., Bradley, K.A., Burke, K.A., Helmick-Rich, J.A.. Greenwood-Ericksen, A., & Hancock, P.A. (2003)*

Moderating Factors in Visual Search: The Role of Ecological Validity *Proceedings* 47th Annual Human Factors and Ergonomics Society *Diaz, D.D., Sims, V.K., Hancock, P.A., Smith, H.S., Ellis, U., Clark, B., & Syshil, D. (2003)*

An Investigation of Individual Differences in Relation to Warning Compliance and User Affect: Preliminary Findings *Proceedings* 47th Annual Human Factors and Ergonomics Society *Weaver, J.L., Helmick, J.A., & Burke, K.A. (2003)*

Application of Human Factors to Homeland Security *Presentation* 49th *Southeastern* Psychological Association, *Hancock, P.A. (2003)*

Application of Fuzzy Signal Detection Theory to Law Enforcement *Poster* 49th *Southeastern* Psychological Association *Szalma, J.L., Hancock, P.A., Weaver, J.L., Mouloua, M., & Stafford, S.C. (2003)*

Application of Fuzzy Signal Detection Theory to Homeland Defense *Presentation* 49th Southeastern Psychological Association *Szalma, J.L., Hancock, P.A., & Mouloua, M. (2003)*

Individual and Dual Task Performance Decrements as a Function of Stressor Exposure *Poster* 49th Southeastern Psychological Association *Bradley, K.A., Greenwood-Erickson, A., Weaver, J.L., Helmick, J.A., Burke, K.A., & Hancock, P.A. (2003)*

Individual and Dual Task Performance Decrements as a Function of Stressor Exposure *Abstract* 49th Southeastern Psychological Association *Bradley, K.A., Greenwood-Erickson, A., Weaver, J.L., Helmick, J.A., Burke, K.A., & Hancock, P.A. (2003)*

Adaptive Experts and Automaticity: Explanations of Failing Under Pressure *Abstract* Cashiers *Ward, P., Szalma, J.L. , and Hancock, P.A. (2003)*

A Research Methodology and Initial Findings: Studying Combat Readiness and Stress within Teams *Presentation* American Psychological Association Mid-Year *Burke, C.S., Bowers, C., Salas, E., Asberg, K., Priest, H., & Milham, L. (2003)*

Facilitated Work

Demonstrating a Methodology for Assessing Cognitive Capacities Across Sensory Systems: M-SWAP *Proceedings* 47th Annual Human Factors and

Ergonomics Society *Brill, J.C., Mouloua, M., Hancock, P.A., Gilson, R.D., & Kennedy, R.S. (2003)*

The Effect of Affect: The Hedenomic evaluation Human Computer Interaction *Proceedings* 47th Annual Human Factors and Ergonomics Society *Murphy, L., Stanney, K., & Hancock, P.A.*

M-SWAP: Online Workload Assessment *Abstract* 12th International Symposium on Aviation Psychology *Hancock, P.A., Brill, J.C., Mouloua, M., Gilson, R.D. & Kennedy, R.S. (2003)*

Skills, Rules and Knowledge: An Experimental Test of Performance Decrements as a Function of Stressor Exposure *Abstract* 12th International Symposium on Aviation Psychology *Weaver, J.L., Bradley, K.A. & Hancock, P.A. (2003)*

Catholic University (Raja Parasuraman)

Supported Work

Effects of Task Duration and Type of Automation Support on Human Performance and Stress in a Simulated Battlefield Engagement Task *Proceedings* 47th Annual Human Factors and Ergonomics Society *McGarry, K., Rovira, E., & Parasuraman, R. (2003)*

University of Minnesota (Wayne Harris)

Supported Work

Field Assessment of Cognitive Performance under Stress *Proceedings* 47th Annual Human Factors and Ergonomics Society *Harris, W.C., & Hancock, P.A. (2003)*

Kansas State (Kip Smith)

Supported Work

The Effects of Trust, Communication and Proximity on Team Problem Solving *Proceedings* 47th Annual Human Factors and Ergonomics Society *Pruitt, N. Bolton, J., Kearney, A., Swan, B., & Smith, K. (2003)*

Communication, Proximity, and Trust in Distributed Command and Control *Proceedings* 47th Annual Human Factors and Ergonomics Society *Swan, B., Younger, B., Smith, J., Toll, M., & Smith, K. (2003)*

Distributed Teams and Revealed Trust *Proceedings* 47th Annual Human Factors and Ergonomics Society *Bloomberg, S., Hawkins, S., Davidson, R., Swan, B., & Smith, K. (2003)*

Beyond Milgram: An experimental study of leader presence *Proceedings* 47th Annual Human Factors and Ergonomics Society *Pangburn, K., Freund, S., Pangburn, H., & Smith, K. (2003)*

Initial Experiments on Cultural Factors that May Mitigate the Adverse Effect of Distrust on Team Cognitive Performance under Live Fire *Paper* Human Performance, Situation Awareness and Automation Conference *Smith, K. (2003)*

Revealed Trust and Team Performance *Paper* North American Meeting of the Economic Science Associations *Smith, K. (2003)*

Progress Report – Smith, K. 01/2003 – 06/2003

Progress Report – Smith, K. 06/2003

Progress Report – Smith, K. 3rd Quarter 2003

Progress Report – Smith, K. 07/2003 – 09/2003

University of Cincinnati (Joel Warm)

Supported Work

Cerebral Vascularity and Performance on an Abbreviated Vigilance Task *Proceedings* 47th Annual Human Factors and Ergonomics Society *Hollander, T. D., Helton, W. S., Tripp, L.D., Parsons, K., Warm, J. S., Matthews, G., Dember, W. N., and Parasuraman, R.*

Challenges to the Mindlessness Model of Vigilance Through Signal Regularity *Proceedings* 47th Annual Human Factors and Ergonomics Society *Helton, W. S., Hollander, T. D., Warm, J. S., Matthews, G., Dember, W. N., Wallaart, M., Beauchamp, G., Parasuraman, R. (2003)*

Facilitated Work

Effects of Transitions in Signal Salience on Vigilance Performance and Stress *Paper* 95th Annual Southern Society of Philosophy and Psychology *Helton, W. S., Shaw, T. H., Warm, J. S., Matthews, G., Dember W. N., & Hancock, P. A. (2003)*

Progress Report - Warm, J. 01/2003

Progress Report - Warm, J. 04/2003

Progress Report - Warm, J. 07/2003

Progress Report - Warm, J. 09/2003

THEOR. ISSUES IN ERGON. SCI., 2003, VOL. 44, NOS. 1–2, 238–249 Taylor & Francis Taylor & Francis Group

The future of neuroergonomics

P. A. HANCOCK and J. L. SZALMA*

Department of Psychology, and Institute for Simulation and Training, University of Central Florida, PO Box 161390, Orlando, FL 32816-1390, USA

Keywords: Brain and behaviour; ergonomics; ethics; human factors; intentionality; neuroscience.

This paper addresses the theoretical, the philosophical, and the ethical considerations associated with the advent and future of neuroergonomics. These issues will demand increasing attention as knowledge of the linkage of brain function to technology-based action improves from its current coarse-grained level to a more fine-grained understanding. These developments promise to open extraordinary opportunities for improved human–machine and human–human interaction, and represent the next major step in human–machine evolution. The social and psychological implications of these changes, however, must be considered if abuse of this conception is to be circumvented.

1. Introduction

Laid before us now is the vision of a brave new world in which technology promises to grant and fulfill our every momentary wish as effortlessly as thought itself. Set against this Arcadian vision is an ever-present nightmare of machine domination in which we forfeit the very humanity by which we define ourselves. Between these two extremes lies the spectrum of future realities. Redolent of modern science fiction literature, such vistas may be rendered one step closer to reality by the birth of neuroergonomics (Parasuraman 2003). We are firm believers in the sequential dependency of history, in that actions taken now at the genesis of the next phase in human–machine interaction circumscribe what the possible futures will be. Consequently, in the present paper we offer more than a commentary on the papers in the two special issues of this journal on neuroergonomics. Rather, using knowledge garnered from these works, we wish to evaluate what directions neuroergonomics might take and what issues need be considered as such progress is engaged.

Consider the development of perceptual-motor skill. We have all seen the haphazard, effortful and sometimes comic actions of the naïve performer. Often the beginner approaching a new skill will make large, radical, excursive motions in an attempt to capture even the semblance of the required movement pattern. In stark contrast are the polished, elegant actions of the highly skilled individual which hardly seem to involve any effort at all. It is our view that the state of current human interaction with technology is more akin to the activities of the naïve beginner as compared to the polished response of the expert. This is largely because current human interaction with technology has to proceed through highly limited input–output capabilities curtailed by the qualities of perceptual processes and the response capacities of the motor system. That the contemporary inelegance of

* Author for correspondence. e-mail: jszalma@pegasus.cc.ucf.edu

Theoretical Issues in Ergonomics Science ISSN 1463–922X print/ISSN 1464–536X online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI 10.1080/1463922021000020927

interaction is due to the miseries of current interface design is evident. However, it is also intrinsically limited by this traditional mode of interaction with the world. Eyes, ears and skin for input and fingers, toes and voice for output are very limited ways for the brain to convert imagination to reality. We see neuroergonomics as the next, iterative step in expanding communication channels between humans and technology, a progress that has begun to accelerate as knowledge in the neurosciences has advanced beyond its first primitive stages.

While many early forms of tools were custom-made, the explosion of the 19th century Industrial Revolution created a one-size fits all mentality which percolated beyond the workplace into the manufactured artifact itself. The advent of the information age belied this nomothetic domination and we began to see adaptation as an important design issue in human–machine systems (Hancock and Chignell 1987, Rouse 1988). Such adaptation was a general property and did not rely on the presence of a single particular individual. For example, the Pilot's Associate project, an adaptive expert system for fighter pilots (Hammer and Small 1995), was designed to work generally rather than with a specific individual. We have now moved to a situation in which we have to pay attention to individuals and design must be flexible enough not simply to present a restricted smorgasbord of choices, but rather truly adapt to the singular individual character of the current user. But how is this to be done?

In these special issues are brought together discussions of diverse applications and methods of neuroscience techniques that promise to provide answers to the above question. Each article shares a common theme of relating brain activity to information processing and, in doing so, showing both the potential power of neuroergonomics and its current limitations. For the future of neuroergonomics, we see the following major issues to be addressed (although we do not claim this is an exhaustive list):

(1) The specificity of brain state assessment (brain–cognition link).
    (a) Neural structure, psychological constructs and reductionism.
    (b) From neurons to intention.
(2) Synergy of neuroergonomics and cognitive neuroscience.
(3) The potential of neuroergonomics.
(4) Ethical and philosophical concerns.
    (a) Freedom.
    (b) Privacy and information ownership.
    (c) The fracture of unitary consciousness

2. Brain–cognition link: the measurement problem

Currently, we have only very gross measures of the functional brain status of any specific individual. Measures of mental or cognitive workload take output either from the motor system via primary and secondary task performance, from the central and peripheral nervous system (as physiological reflections) or from subjective report. These indicators were perhaps the first used to capture the dynamics of human–machine interaction. However, today these measures may appear 'antiquated' in light of the exciting, multi-dimensional pictures provided by the various new brain-imaging techniques, but the fundamental problem remains the same as it has been throughout the history of psychology—how do we connect thought to action?

Primarily, this is a problem of measurement, which in turn depends upon one's theories of neuropsychology and its relation to neurophysiology. If one asserts that psychological theories exist only because we cannot adequately specify neurological function, then cognitive neuroscience and neuroergonomics become fields that, ultimately, replace psychology and human factors as information processing theories are rendered obsolete by increasing specification of the brain–cognition link. This amounts to a monistic, reductionist approach in which psychological processes are reduced to physical ones when the latter are sufficiently understood.

Alternatively, one could take a functionalist approach more typical of cognitive psychology (see Parasuraman 2003), in which the knowledge of neurological structures does not largely constrain understanding the structure and function of cognitive processes. A less severe form of this argument is a dualist concept, that distinct physical and psychological constructs are both necessary for a complete understanding of cognition (and human behaviour in general). In this view, one might posit that there are 'emergent features' at the level of cognition that are lost when they are reduced to their physiological correlates, but that understanding neurological structures provide one of many windows into cognition.

Regardless of which of these philosophical positions one adopts, the fundamental problem of establishing good neurological measures of cognitive processes remains. This, in turn, depends on the quality of our psychological theories, since psychological constructs lend meaning to the neurological data. For example, to link EEG patterns to cognitive processes the latter must be adequately specified, and this challenge was elucidated by the authors of some of the preceding papers. For example, Scerbo *et al.* (2003) argued that EEG indices of workload must be sensitive and diagnostic for a given task, and that the engagement index they derived from EEG data must be extended beyond specific tasks such as tracking and vigilance. The broader issue is how we know that the new EEG indices actually reflect mental workload. As Baldwin (2003) pointed out, one limitation of performance-based measures of workload is that performance reflects more than one process and is, therefore, not a unique measure of mental workload (see also Gopher and Donchin 1986, Meshkati *et al.* 1990). Similarly, a challenge for neuroergonomics is to establish that specific neurological indices reflect mental workload and no other process, an exclusivity criterion that is always difficult to establish in any multifactorial situation. A complicating factor is the possibility of different patterns of workload/performance associations and dissociations (Yeh and Wickens 1988, Hancock 1996, Parasuraman and Hancock 2001). These have been observed with subjective measures of workload and they may persist as relations between performance and neurological indices of workload are established across different tasks. This point was obliquely touched upon by Scerbo *et al.* (2003) when they discussed cases where performance and mental workload were not correlated. They discussed this issue in the context of the difficulty of validating EEG against behaviour, but in addition we suggest that such relations may enhance the validity of EEG by showing (potentially) that specific patterns of associations and dissociations between EEG and performance are linked to particular task categories. A related issue concerns the inconsistency between subjective and physiological measures that have also often been observed. This issue presents important challenges for the advancement of neuroergonomic approaches to workload and, in addition, to similar problems in establishing neurological indices of all forms of stress-state. Like fatigue, workload and stress

are multidimensional constructs and neurological indices of them must be sensitive to this multidimensional facet (see Hancock and Desmond 2001).

Neuroergonomics can contribute to a more accurate assessment of all multidimensional cognitive states. As Scerbo *et al.* (2003) and Sarter and Sarter (2003) clearly articulated, neural measures have the potential to allow assessment of important constructs in cases where performance and subjective measures are either impractical or impossible. We would add, however, that a multi-assessment approach is vital here. Neural measures should be explored, but in practice they should be used whenever possible in combination with other measures (i.e. performance; subjective response), so that design decisions are not based upon only one order of data alone.

2.1. *Neural structure, psychological constructs and reductionism*

The development of good neurological measures of cognitive processes is limited by the fact that many of our psychological constructs (e.g. mental resources) remain poorly defined. Since the power of neuroergonomics will come from its validity for measurement of cognitive activity, established by correlating physiological matrices to cognitive measures, the validity of neuroergonomics will rest to some degree on the validity of our psychological models (and measures) of cognition and human performance. A footnote by Hettinger *et al.* (2003) nicely illustrates this concern. They pointed out that the ability to develop neuroadaptive interfaces depends on our ability to identify the relevant cognitive and emotional states associated with a given domain or set of tasks. As is clear from each of the papers in these special issues, the development of neuroergonomics must be in tandem with not only cognitive neuroscience but also cognitive psychology.

We accept the difficulty in establishing that a physiological index really reflects a specific aspect of cognition. However, we should note that this is a problem for *all* measures of human behaviour and mental processing. For instance, the problem of valid measurement of workload is as difficult to establish for subjective report as it is for psychophysiological measures. Sanderson *et al.* (2003) faced a similar limitation in that neural imaging techniques such as Steady State Probe Topography (SSPT) cannot, by themselves, identify displays (or any interface, for that matter) as 'good' or 'bad'. This is again due to the constraint we have articulated, which is that there is no simple mapping between cortical activity and cognitive processes. The notion that more is gained by replacing constructs like workload and situation awareness with a detailed neural identification of the cognitive processes underlying the constructs suffers the misguided assumption that reduction ultimately produces 'the' explanation for a phenomenon. A strategy of endless reduction can only lead ultimately to the level of quantum mechanics, which itself is struggling to link concepts to an understanding of consciousness.

The issue of reductionism is specifically addressed here by Sarter and Sarter (2003). They argued that for neuroergonomics to fully benefit from the approaches of cognitive neuroscience, researchers and practitioners in cognitive ergonomics must be willing to embrace a reductionistic approach. This would involve reducing higher level, multidimensional psychological constructs (they use examples of decision-making, situation awareness and mental workload) to underlying cognitive activity closely tied to brain activity. They noted, however, that our understanding of psychological constructs such as mental workload requires psychological analysis that is not necessarily enhanced by correlating performance with psychophysiological data. The latter have value, but do not replace the former. This is an important

point to consider as the field of neuroergonomics develops. This approach promises to yield great benefits, but one must not fall into the trap that reducing explanation from one level to a more 'basic' level will obviate the need for analysis at the higher level. Simultaneous exploration of cognitive processes at the neural and gross psychological levels must be maintained. Indeed, psychological concepts such as 'resources' should benefit from this fusion (Szalma and Hancock 2002).

2.2. *From neurons to intention*

The problem of intention pertains to these difficult measurement issues. While neural activity will, with greater precision, be linked to specific cognitive processes, it is less clear how summed knowledge of multiple neuronal states will provide information regarding intention. Indeed, is it reasonable to talk of these two spheres of discourse in the same manner? Motivational constructs are especially difficult to specify at the psychological level and this intrinsically limits the potential for specific neural processes to indicate specific motivational states. Beyond the problem of establishing a valid definition of attributes such as motivation, there is also the issue of the locus or source of intention. For instance, it is possible that the brain mechanisms that react to extrinsic motivation are distinct from those that control intrinsic motivation. Efforts to define motivational states in terms of highly specified cognitive models (e.g. semantic networks) will aid in a partial resolution of the intention issue, but the ultimate source of intention (the homunculus problem) will not likely be resolved soon by neuroscience.

3. Synergy with cognitive neuroscience

Sarter and Sarter (2003) correctly noted that neuroergonomics will have to adopt characteristics of cognitive neuroscience, since the goals for these fields are so closely related. Cognitive neuroscience seeks the discovery of brain mechanisms mediating complex cognitive activity, while neuroergonomics seeks to understand how those same brain mechanisms are involved in human performance in interaction with technology. We appear to be at a point similar to that of Human Factors in the 1940s and 1950s, as it adopted the approaches of experimental psychology, particularly perceptual psychology, in the investigation of human–machine interaction. As Hettinger *et al.* (2003) argued in their article in this special issue, a strength of neuroergonomics is its potential to link diverse lines of research in human factors, experimental psychology and neuroscience, and it can be used to explore how facets studied separately by other fields (e.g. perception, cognition, motivation, emotion) interact to impact (and produce) behaviour.

The importance of cognitive neuroscience to neuroergonomics argued so cogently by Sarter and Sarter (2003) was underscored by the complexity of the results and the difficulty in data interpretation reported by Sanderson *et al.* (2003) and Scerbo *et al.* (2003). Neuroscience techniques are not a way around the limits of behavioural measures. As discussed earlier, they will work only to the extent that we have good psychological theory to define information processing structures to which brain function can be related.

4. The potential of neuroergonomics

One of the prime forms of application of neuroergonomics is to the creation and operation of adaptive systems (Hancock and Chignell 1987). As discussed by Scerbo *et al.* (2003) and Hettinger *et al.* (2003), there is more to be done before the full

potential of these systems can be realized. Here, we address some of the issues for future research in adaptive systems as well as other areas in which neuroergonomics can be employed.

The potential for adaptive interfaces to improve the human–machine interaction is clear. Hettinger *et al.* (2003) discussed how these systems could be used to modify the presentation of information to best fit the person's mental state. For instance, an adaptive interface based on neural measurement could, if reliable, make an excellent alarm system, as a change in brain state indicating inattentiveness to an environmental event could be detected by the computer, which could sound the alarm. In addition, neuroergonomically designed systems could allow thought to be directly linked to system action, so that, for example, individuals with physical impairments could be trained to perform neural control of a computer-based system, opening up new activities for them (see Parasuraman 2003).

The example of adaptive interface design for the physically impaired illustrates that the possible applications of neuroergonomics extend beyond consideration of cognition and include physical activity as well. For instance, Karwowski *et al.* (2003) described studies on the brain-action relations in muscle control. From these studies, we can envision the development of adaptive aided action that parallels adaptive automation in which a computer-based system could sense muscular limitations of an operator (e.g. fatigue) and adapt human input requirements accordingly. Karwowski *et al.* (2003) discussed the potential for Physical Neuroergonomics to alleviate musculoskeletal injury, but the implications for improving human–environment interaction go further than prevention alone. Human physical capability may be extended in both strength and speed such that, with sufficient gain, minute muscular responses could produce physical activity beyond the limits of human-range (i.e. unaided action), as discussed by Hancock (1997). In this case, thoughts would directly produce movement of the self or other objects. Returning to the earlier example of design for the physically impaired, these individuals would gain the ability to execute physical behaviours directly from the brain, thereby expanding their ability to act on their environments beyond computer-based information processing tasks (cognition) to any physical task currently beyond their action capabilities.

Hettinger *et al.* (2003) cast the problem of human–machine interaction in terms of dysfunctional communication which may be solved by enhancing communication using measures of neural activity. This offers a promising approach to addressing what they referred to as asymmetrical communication between humans and machines, which they likened to 'dysfunctional interpersonal communication' between people. For example, they noted that a computer is ill equipped to recognize the changes in a human's cognitive or emotional state and to subsequently modify its behaviour to accommodate those changes. They argued that neuroadaptive interfaces could address this problem by allowing more precise monitoring of the operator's mental state. Note, however, that developing a computer system that can assess a human's cognitive or emotional state makes the machine functionally more 'human'. One issue for future research will be the implications of such a system for function allocation, which has traditionally been based on fundamental differences between humans and machines. As machines take on human abilities (e.g. 'emotional perception'), what will the role for the human be? Indeed, as the synergy between human and machine is increased by direct links between brain activity and machine activity, with an operator's thoughts directly impacting

machine performance, and the machine, in turn, directly influencing brain activity, the boundaries of human identity and consciousness blur, and we consider this in more detail below. It may well become difficult to determine, at a functional level, where the human ends and the machine begins. While this will likely substantially enhance system performance and has extraordinary potential for improving quality of life (e.g. for the physically impaired), it also raises important philosophical issues regarding human identity and changes in the human condition.

Some of the issues which surround the neuroergonomic approach to adaptive automation are similar to those already discussed for general adaptive interfaces. The work of Scerbo *et al.* (2003) also indicates that function allocation will have to be reconsidered with adaptive automation. As they pointed out, negative and positive EEG feedback have different effects on performance. An implication of this, however, is that giving the operator something to do is important for both system performance and operator well-being. While this has been discussed in other contexts (Hancock 1998, Scerbo 1998), the physiological evidence supports the argument that performance suffers when automation leaves the human with little to do (Parasuraman and Riley 1997, Parasuraman and Hancock 2001). Specifically, Scerbo *et al.* (2003) showed that in positive feedback conditions automation is activated when the person is 'disengaged' (i.e. inattentive) and deactivated when the person is attentive and engaged in the task. Negative feedback activates automation when the person's attentiveness increased and automation turned off when attention waned. Across several experiments, performance was better with negative feedback rather than positive. In other words, performance was better when automation was turned off at the point the person's engagement decreased, possibly activating the person to become more involved in the task. Hence, letting the machine take care of things when a person becomes less attentive does not necessarily improve performance.

Findings such as those described above indicate that an important question for a burgeoning neuroergonomics is how (or whether) such an approach to the study and design of human–machine interaction will relieve stress and workload, and whether the heavy loads will be replaced with the burden of boredom. In principle, a computer could detect 'boredom' and respond by providing stimulation that is engaging for the users and captures their interest. If applied properly, this would be an application of neuroergonomics to *hedonomics*, which has been defined by Hancock as *that branch of science which facilitates the pleasant or enjoyable aspects of human–technology interaction*. Hedonomics contrasts with the traditional ergonomic goal of reducing damage, injury, inconvenience and frustration. The primary concern for hedonomics is whether one can design for happiness. Neuroergonomics holds great promise for advancing this effort, as it will permit dynamic adjustments of interfaces to match the emotional state of the person and will permit individual differences in preferences to be designed into the interface. Note that such an application will be difficult, since one is trying to link fine-grained neural indices to cognitive functions that, psychologically, are articulated at a molar level (i.e. what does 'pleasant' or 'enjoyable' mean?). In addition, issues of human identity and intention discussed earlier will need to be addressed to insure that efforts to apply hedonomic principles to interface design are not undone by loss of human identity and autonomy.

One avenue for further research to answer questions regarding stress, workload and human performance is to explore whether the EEG correlates of attention and cognition will persist under different environmental conditions. That is, does

environmental stress (or task stress, for that matter) qualitatively change the EEG pattern, or is the pattern the same, with the relative magnitudes of α-, β- and τ-waves shifting? It may be that a specific EEG pattern is observed when an operator is stressed or overloaded, but such a pattern may be specific to particular domains.

The papers discussed in this special issue address the establishment of brain-action relations (Karwowski *et al.* 2003), of brain-cognition relations and their application to adaptive systems (Hettinger *et al.* 2003, Scerbo *et al.* 2003), and to enhancing our understanding of the mental processes underlying performance on current display designs (Sanderson *et al.* 2003). As neuroergonomics develops, fruitful avenues for research and practice may be found in the other direction—using neurological information to design new interfaces for physical and cognitive tasks. To date, the design of controls and displays has been constrained by limitations of the human sensory and motor systems. In the future, the brain will be added as another physical structure around which ergonomic principles can be applied for interface design. The advancement of neuroergonomics and cognitive neuroscience may yield a complex neuroanthropometry for work design. In addition, with sufficient advances systems will be designed in which thought is immediately transformed into action, allowing smooth, effortless human–machine interfaces. Advances in physical neuroergonomics will be essential in this regard. As Karwowski *et al.* (2003) noted, the two areas of physical and cognitive ergonomics have not been sufficiently integrated, and this integration, while important for ergonomics, is essential for the growth of neuroergonomics if we are to achieve a comprehensive understanding of the dynamics of brain–environment interaction.

5. Ethical and philosophical concerns

A clear conclusion from the papers in these special issues is the potential for neuroergonomics to improve human–technology interaction, thereby improving overall quality of life (see Hancock 1997). However, there is also the potential for negative effects, particularly in issues of freedom, privacy, and the loss of ownership of one's mental state, and it is to these that we now turn.

5.1. *Freedom*

As neuroergonomically designed systems are implemented, concerns over freedom, privacy and property are bound to present problems which exceed the purely technical. In regard to freedom, as we move forward we must insure that in designing these systems we do not, even implicitly, create them so that the human is treated as an alienated extension of the machine (see Fromm 1941). This will involve developing ethical guidelines in the same way as we need to address cloning and human DNA research. In particular, de-centralized organizations in which autonomy is granted to the employees in defining the parameters of their tasks would be most amenable to preserving human control and dignity. Fortunately, the initial outlines of such de-centralization have been developed, both for organizations and society as a whole (Fromm 1955, Illich 1973, Deming 1986, 1993). If we develop the technology without changing our social and organizational structures, individuals who hold power in organizations will, knowingly or unknowingly, relegate their employees to the status of machine parts and treat them as such. The issue will become one of autonomy and authority—who controls an individual's mental state? Does the operator retain control or does an external authority influence a person's mental state via the machine?

5.2. *Privacy and ownership*

Another important ethical issue to be considered is the preservation of privacy. Hettinger *et al.* (2003) are correct in warning about the dangers of invasion of privacy, personality profiling and other negative effects systems designed from neuroergonomic principles could have for human well-being. We must be careful to not surrender too much control (in the form of information) to the owners of neuroergonomically designed technology. In this regard, a related issue is the ownership of the information on neural states obtained from these systems and who has access to that information. Current legal structures allow the organization for which a person works to retain ownership on intellectual property the employee creates at work (and, in some cases, outside of work—e.g. where employees surrender all patent rights to their employer). These structures will have to be changed to reflect the changes in information produced by developments in neuroergonomics. Ownership of neural information regarding mental state should be retained by the individual (in more than a purely legal sense) and access to the information should be highly restricted. In addition, organizational structures should be created that hold accountable those who do have access to such highly sensitive information. These issues may seem far from those typically considered in human factors and ergonomics, but these are the new vistas and concerns resulting from the advent of neuroergonomics. We should not fall prey to the mistake of waiting until technology is implemented before we address the ethical and legal implications of the technology. Now is the time for these philosophical issues to be addressed.

5.3. *Fracturing unitary consciousness*

Throughout history, human beings have been defined by their individuality. The events of life are experienced alone and the quintessential separation between individuals provides much of the motive force of art and philosophy as we seek to understand ourselves and our peers. Knowledge of our own certain and individual demise defines the human condition and characterizes us as a species (Hancock 2002). The advent of neuroergonomics, if its vision is realized, will transform this condition. As we noted earlier, the ways in which the brain can receive information and achieve intention are very coarse compared with what they might be. However, as neuroergonomics becomes more successful, the barrier between the human and the machine promises to dissolve and the issue of human identity noted earlier will come to a head. As that barrier is breached, the consequent separation between human beings and, thus, the unitary nature of consciousness will begin to fracture. As these stages of evolution are achieved the empirical question concerns the nature of consciousness itself. We believe that, like many other vestigial characteristics, unitary consciousness will persist but we will become much more adept at working in the context of intimate familiar groups. Neuroergonomically-endowed individuals will begin to become facile with collective group input and emergent social action as forms of response. We do not envisage the complete separation of carbon and silicon forms of intelligence as postulated by Moravec (1988), nor do we foresee replacement of human intelligence by artificial intelligence as argued by Kurzweil (1999). Although it is probable that we shall continue to use machines to explore environments hazardous to humans, the evolution of full human–machine symbiosis, and by extension more intimate human–human interaction, appears to be the path of the future.

6. Conclusions

Neuroergonomics has the potential to improve not only the performance of human–machine systems, but also to improve the quality of life while enhancing our understanding of brain function. The only clearly negative aspect of developments in neuroergonomics is the possibility that the design and implementation of neuroadaptive systems will occur faster than organizational and social progress necessary to accommodate them. If, instead of focusing solely on the capacities of systems, we devote more effort to the teleology of technology (Hancock 1997), we can use scientific understanding to convert intention to action in a benign and empathic manner. In contrast, neuroergonomics would be a dangerous tool in the hands of unfettered capitalism. That intention and ethics can and should be a part of brain sciences and design is yet another opportunity that the present proposed advance to human–machine partnership offers. We intend to watch such progress with interest.

Acknowledgements

This research was supported by a grant from the Army Research Organization (Grant# DAAD19-01-1-0621), Dr. P. A. Hancock, Principal Investigator. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense or the US Government. The authors wish to thank Dr Sherry Tove, Dr Elmar Schmeisser and Dr Mike Drillings for providing administrative and technical direction for the Grant.

References

Baldwin, C. 2003, Commentary: neuroergonomics of mental workload. New insights from the convergence of brain and behaviour in ergonomics research, *Theoretical Issues in Ergonomic Science*, **4,** 132–141.

Deming, E. W. 1986, *Out of the crisis* (Cambridge, MA: MIT).

Deming, E. W. 1993, *The new economics for industry, government, education* (Cambridge, MA: MIT).

Fromm, E. 1941, *Escape from Freedom* (New York: Holt, Rinehart and Winston).

Fromm, E. 1955, *The sane society* (New York: Rinehart).

Gopher, D. and Donchin, E. 1986, Workload: an examination of the concept, in K. R. Boff, L. Kaufman and J. P. Thomas (eds), *Handbook of perception and human performance* (New York: Wiley), 41:1–41:49.

Hammer, M. J. and Small, R. L. 1995, An intelligent interface in an associate system, in W. Rouse (ed.), *Human/technology interaction in complex systems*, Vol. 7 (Greenwich, CT: JAI Press), 1–44.

Hancock, P. A. 1996, Effects of control order, augmented feedback, input device, and practice on tracking performance and perceived workload, *Ergonomics*, **39,** 1146–1162.

Hancock, P. A. 1997, *Essays on the future of human-machine systems* (Eden Prairie, MN: BANTA Information Services Group).

Hancock, P. A. 1998, The price of freedom, *Proceedings of the Human Factors and Ergonomics Society*, **42,** 1577–1578.

Hancock, P. A. 2002, The time of your life, *Kronoscope*, in press.

Hancock, P. A. and Chignell, M. H. 1987, Adaptive control in human-machine systems, in: P. A. Hancock (ed.), *Human factors psychology* (Amsterdam: North-Holland), 305–345.

Hancock, P. A. and Desmond, P. A. 2001, *Stress, workload and fatigue* (Mahwah, NJ: Erlbaum).

Hettinger, L. J., Branco, P., Encarnacao, M. and Bonato, P. 2003, Neuroadaptive technologies: applying neuroergonomics to the design of advanced interfaces, *Theoretical Issues in Ergonomic Science*, **4,** 220–237.

Illich, I. 1973, *Tools for conviviality* (New York: Harper & Row).

KARWOWSKI, W., SIEMIONOW, W. and GIELO-PERCZAK, K. 2003, Physical neuroergonomics: the human brain in control of physical work activities, *Theoretical Issues in Ergonomic Science*, **4,** 175–199.

KURZWEIL, R. 1999, *The age of spiritual machines: When computers exceed human intelligence* (New York: Viking).

MESHKATI, N., HANCOCK, P. A., RAHIMI, S. and DAWES, M. 1990, Techniques of mental workload assessment, in J. R. Wilson and E. N. Corlett (eds), Evaluation of human work: A practical ergonomics methodology (London: Taylor & Francis), 749–782.

MORAVEC, H. 1988, *Mind children: The future of robot and human intelligence* (Boston, MA: Harvard University Press).

PARASURAMAN, R. 2003, Neuroergonomics: research and practice, *Theoretical Issues in Ergonomic Science*, **4,** 5–20.

PARASURAMAN, R. and HANCOCK, P. A. 2001, Adaptive control of mental workload, in P.A. Hancock and P. A. Desmond (eds), *Stress, workload, and fatigue* (Mahwah, NJ: Erlbaum), 305–320.

PARASURAMAN, R. and RILEY, V. 1997, Humans and automation: use, misuse, disuse, abuse, *Human Factors*, **39,** 230–253.

ROUSE, W. B. 1988, Adaptive aiding for human/computer control, *Human Factors*, **30,** 431–438.

SANDERSON, P., PIPINGAS, A., DANIELI, F. and SILBERSTEIN, R. 2003, Process monitoring using a configural display: human performance, self-report, and neuroimaging, *Theoretical Issues in Ergonomic Science*, **4,** 151–174.

SARTER, N. and SARTER, M. 2003, Neuroergonomics: ergonomics or cognitive science? *Theoretical Issues in Ergonomic Science*, **4,** 142–150.

SCERBO, M. W. 1998, Fifty years of vigilance research: where are we now? *Proceedings of the Human Factors and Ergonomics Society*, **42,** 1575–1576.

SCERBO, M. W., FREEMAN, F. G. and MIKULKA, P. J. 2003, A brain-based system for adaptive automation, *Theoretical Issues in Ergonomic Science*, **4,** 200–219.

SZALMA, J. L. and HANCOCK, P. A. 2002, On mental resources and performance under stress, Unpublished white paper, University of Central Florida.

YEH, Y. and WICKENS, C. D. 1988, Dissociation of performance and subjective workload, *Human Factors*, **30,** 111–120.

About the authors

Peter Hancock is Provost Distinguished Research Professor in the Department of Psychology and at the Institute for Simulation and Training at the University of Central Florida. In his previous appointment, he founded and was the Director of the Human Factors Research Laboratory at the University of Minnesota. At Minnesota, he held appointments as Full Professor in the Departments of Computer Science and Electrical Engineering, Mechanical Engineering, Psychology and Kinesiology as well as at the Cognitive Science Center and the Center on Aging Research. He is the author of over 200 refereed scientific articles and publications as well as editing numerous books including: *Human Performance and Ergonomics* in the Handbook of Perception and Cognition series (Academic Press, 1999) and *Stress, Workload, and Fatigue* (Lawrence Erlbaum, 2001). He is the author of the 1997 book, *Essays on the Future of Human–Machine Systems*. In 1999, he was the *Arnold Small Lecturer* of the Human Factors and Ergonomics Society and, in 2000, he was awarded the *Sir Frederic Bartlett Medal* of the Ergonomics Society of Great Britain for lifetime achievement. He was the Keynote Speaker for the International Ergonomics Association and the Human Factors and Ergonomics Society at the 2000 combined meeting in San Diego. In 2001, he won the *Franklin V. Taylor Award* of the American Psychological Association as well as the *Liberty Mutual Prize for Occupational Safety and Ergonomics* from the International Ergonomics Association. He was awarded a Doctor of Science (DSc) degree from Loughborough University in December 2001. In 2002, he was awarded the *Jastrzebowski Medal* of the Polish Ergonomics Society for contributions to world ergonomics and in the same year was named a Fellow of the Ergonomics Society of Great Britain. His current experimental work concerns the evaluation of behavioural response to high-stress conditions. His theoretical work concerns human relations with technology and the possible futures of this symbiosis. He is a *Fellow* of and past President

of the Human Factors and Ergonomics Society. He collects and studies antique maps and is a committed Ricardian.

James Szalma is a post-doctoral fellow in the psychology department and the Institute for Simulation and Training at the University of Central Florida. His research efforts have been focused on the performance, workload and stress of sustained attention, and his current interests include workload and stress, motivational issues in human performance and individual differences in performance. He received his PhD in 1999 from the University of Cincinnati.

Copyright © 2003 EBSCO Publishing

OPUS

Operator Performance Under Stress



University of Central Florida

Kansas State University

Catholic University

University of Minnesota

MURI

MIT

INSTITUTE for SIMULATION & TRAINING

THE CATHOLIC UNIVERSITY OF AMERICA

Multi-Disciplinary University Research Initiative

MIT²

THE CATHOLIC UNIVERSITY OF AMERICA

INSTITUTE for SIMULATION & TRAINING

Operator Performance Under Stress

MURI

MIT²

The Catholic University Of America

Institute for Simulation & Training

University of Central Florida	Catholic University	University of Minnesota	Kansas State University
MIT²	**IST**		
Time Distortion	Teams Under Stress	Brain States & Stress	Physical & Mental Workload	Decision-Making
Interfaces for Stress	Stress & Team Cognition	Cognitive Workload	Physical Evaluation	Distrust
Visual Search	Low Cost Training			
Delta Force	Flashpoint			
Viking		Viking	Viking	Biomedicine

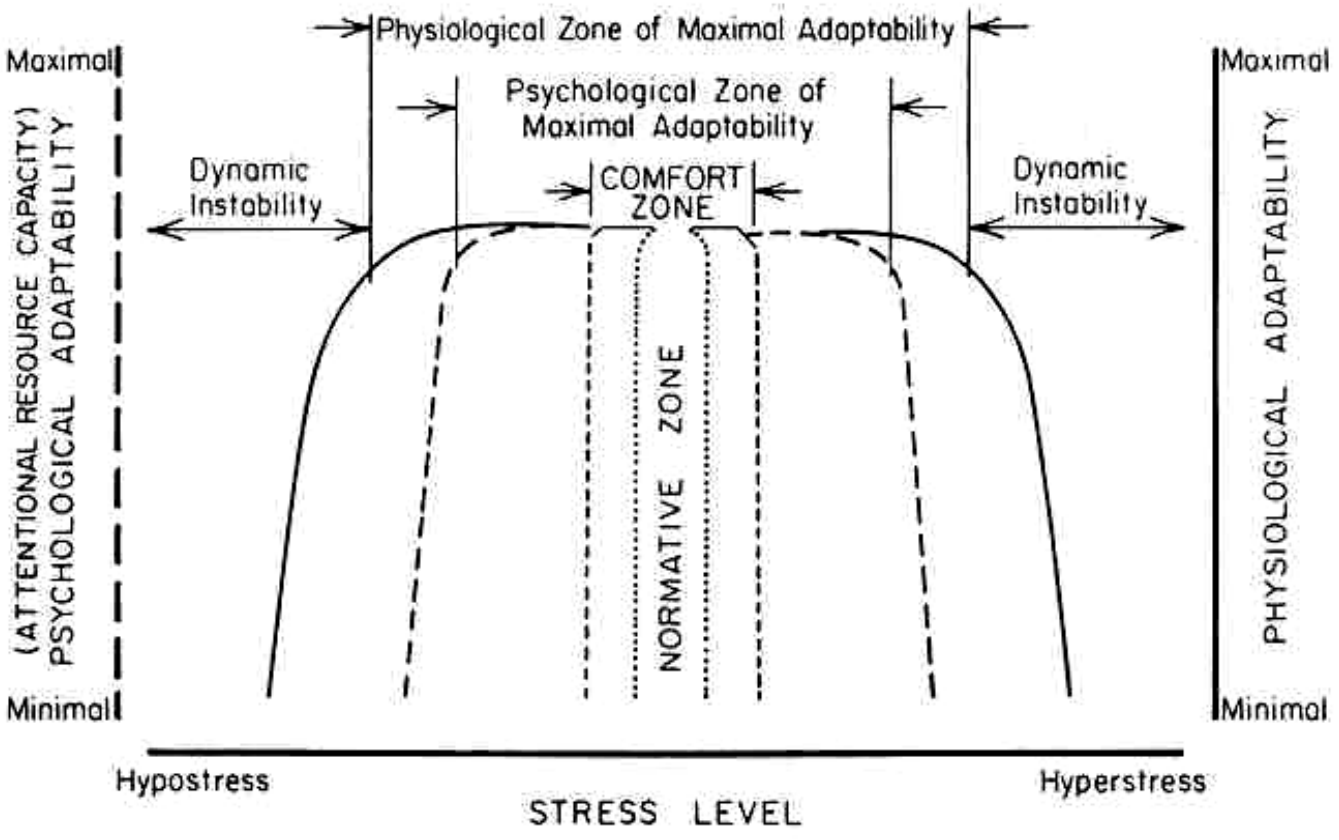
New Affiliated Universities

University of Cincinnati
(Stress Transitions – High/Low Workload)

University of Toronto
(Wearable Computers for Combat)

Florida Institute of Technology
(Adaptive Interfaces for Stress)

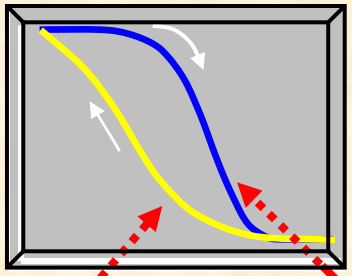
Stress and Performance Theory

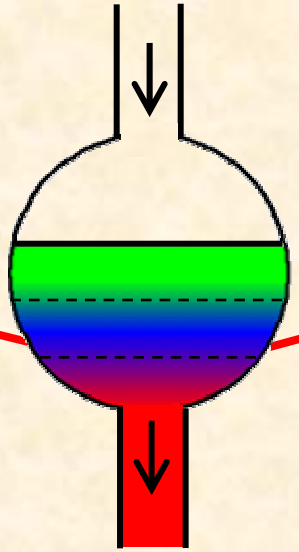


Hancock, P.A., & Warm, J.S., (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.

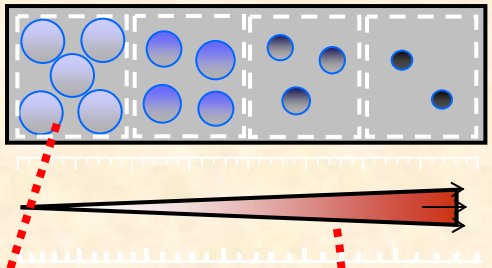
Attentional Depletion/Repletion

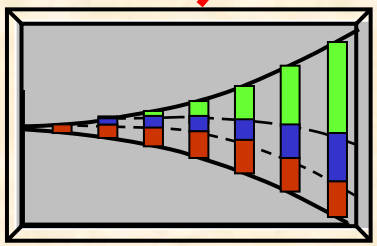
Hystersis
Failure and Recovery



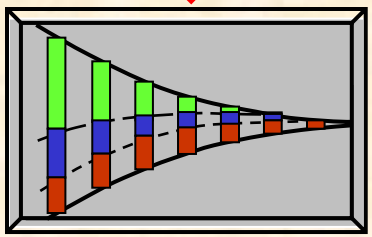


Stress
Distorted Perceptual Space-Time

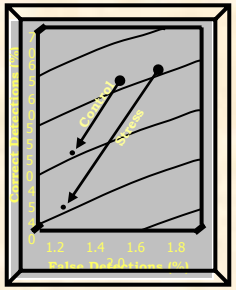




SKR Recovery



Stress Depletion

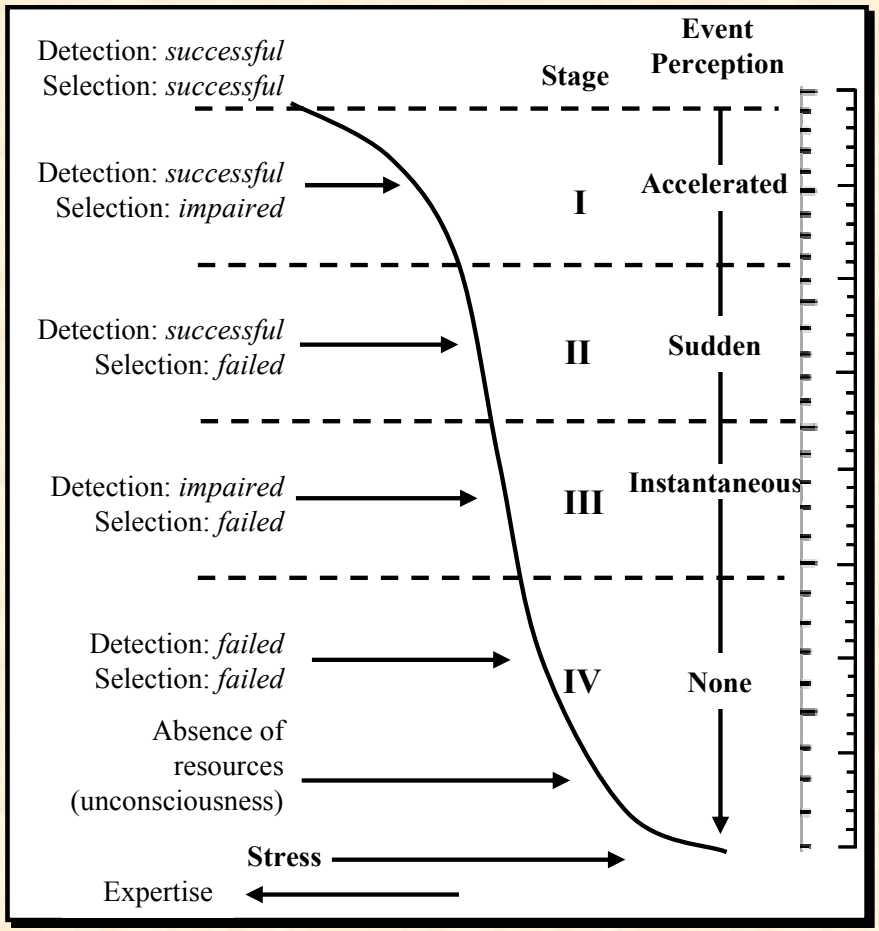


Stress and SDT



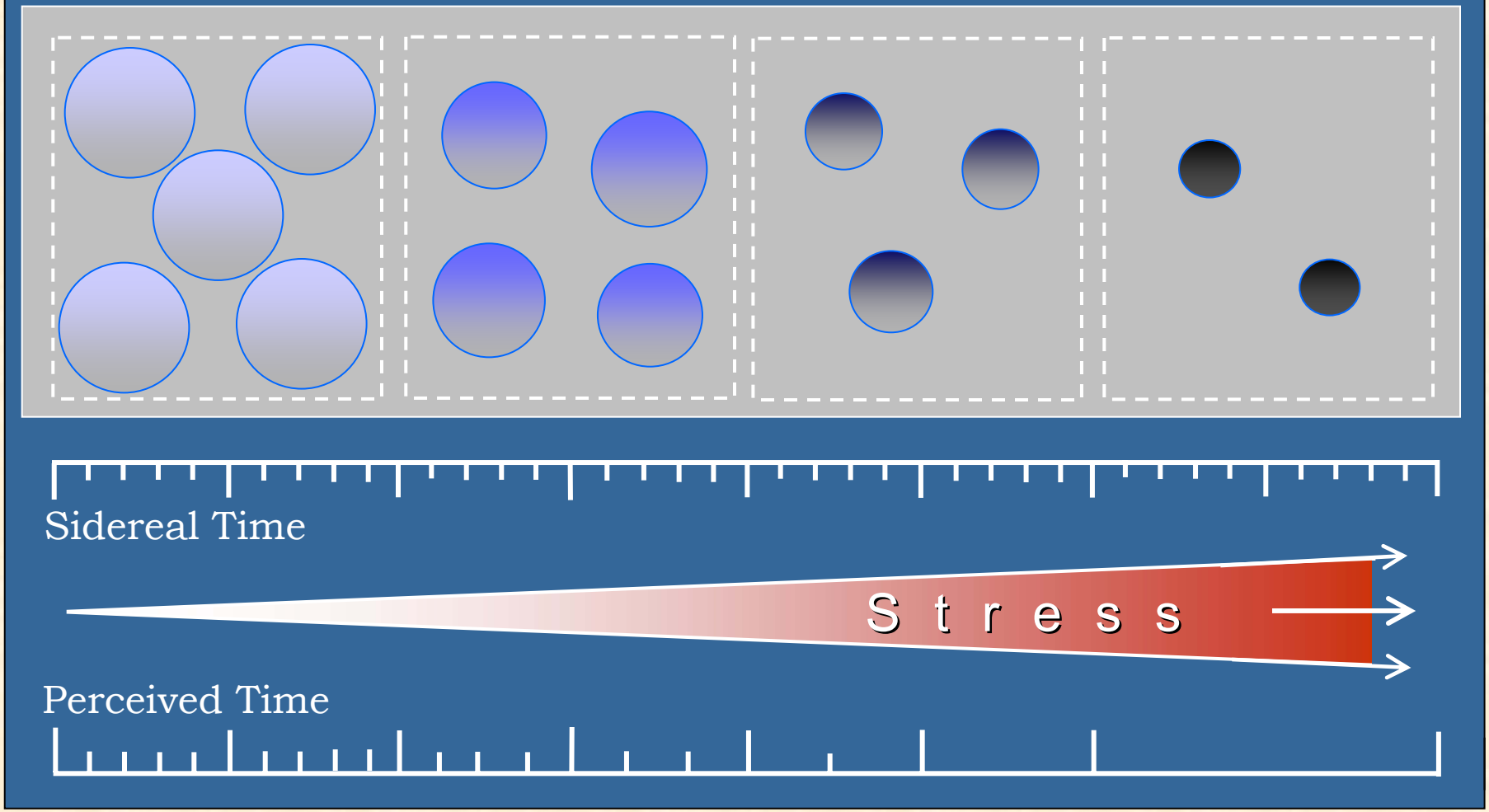
Time Distortion

Stress and Performance Theory



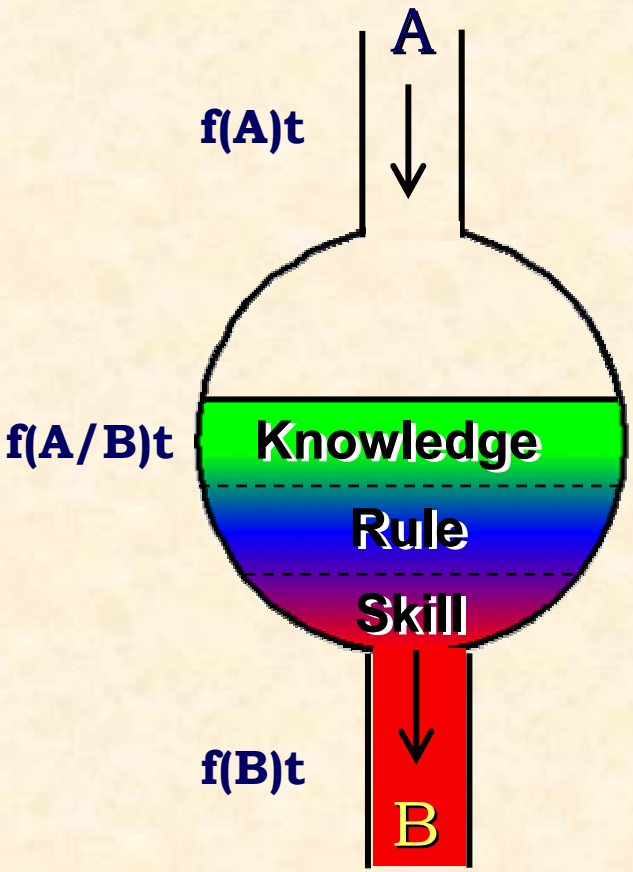
Hancock, P.A., & Weaver, J.L. (2001). Temporal distortions under extreme stress. Theoretical Issues in Ergonomics, VOL.

Stress Distorts Perceptual Space/Time



Stress and Attention

Resource Depletion or Resource Replacement?

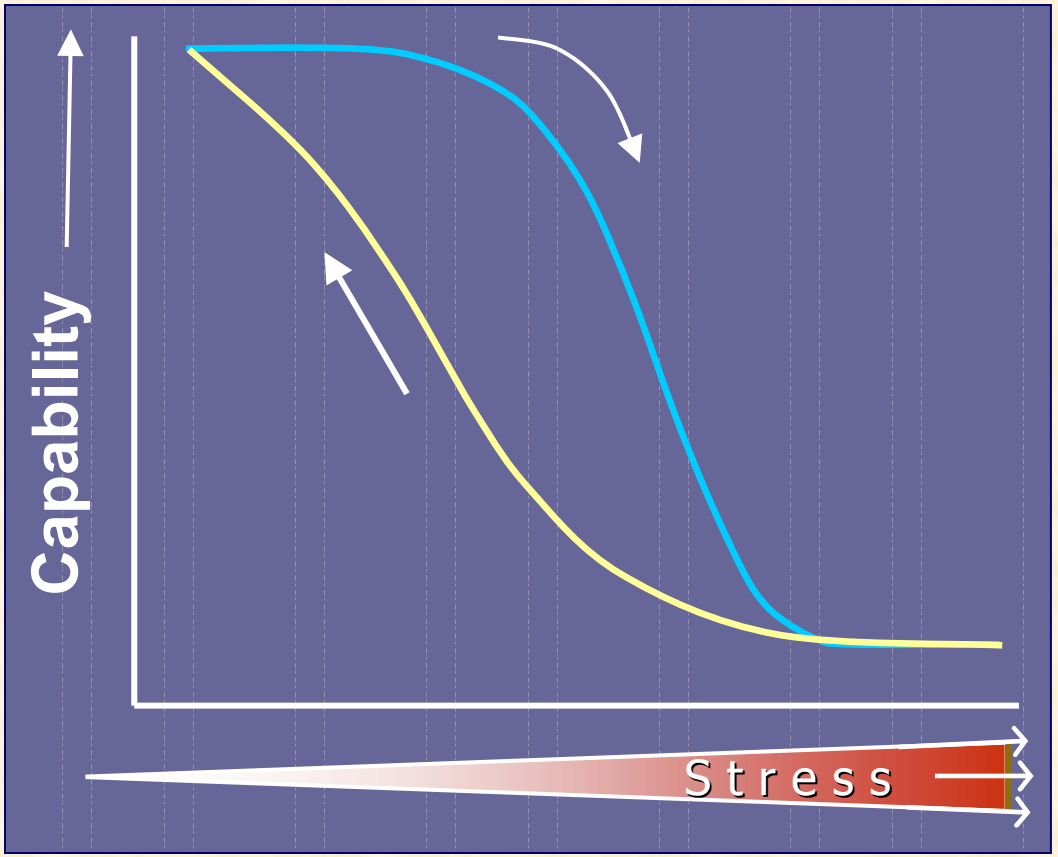


Unexplored aspects of resource change include:

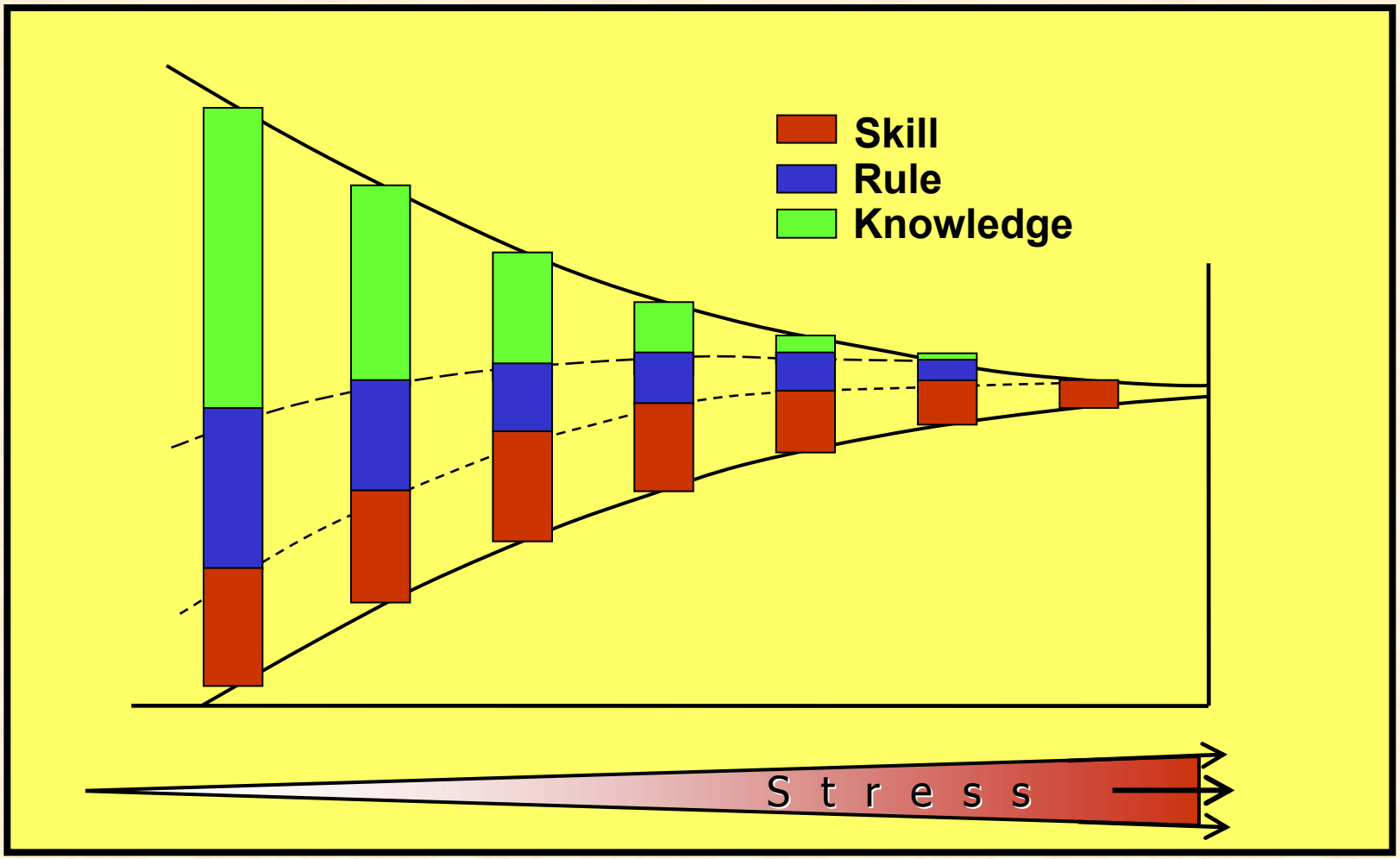
1) Depletion vs. Replenishment
2) Differential Resource Quality
3) Thresholds for Resource Depletion
4) Hystersis in Loss and Recovery

Hysteresis: Stress-Induced Failure and Recovery

Is the way in which the Individual recovers symmetrical with the way the individual fails?



Load-Shedding and Attentional Strategy

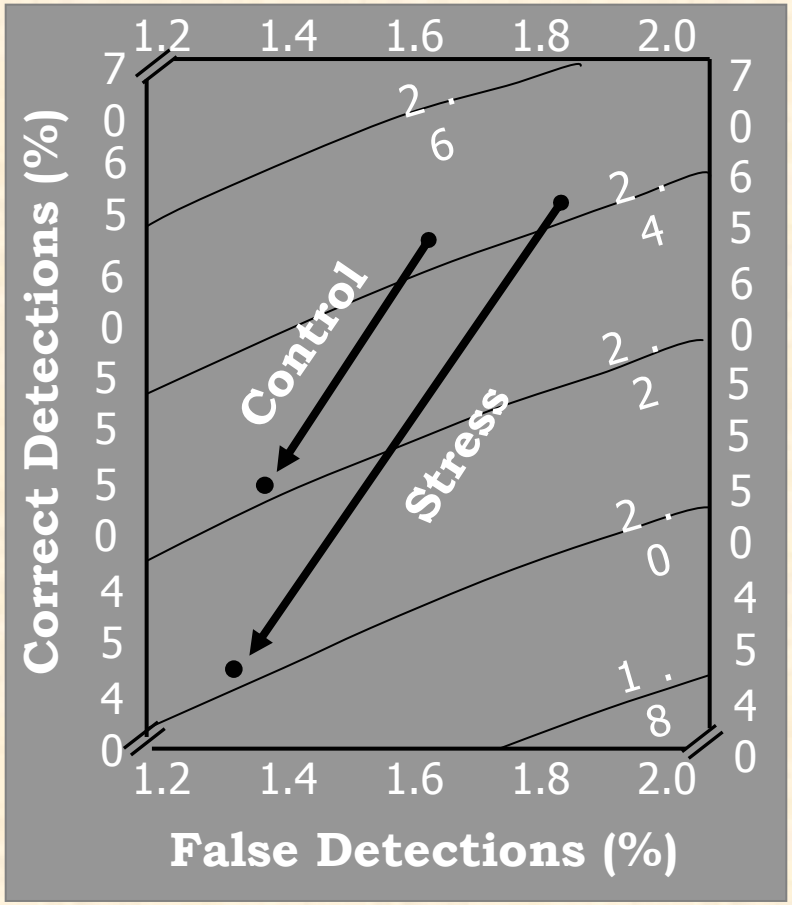


Stress Recovery



Stress and SDT

- How Does Stress Impact d' and B?
- Stress Reduces Positive Responses?
- Fuzzy Signal Detection under Stress?



Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, 42, 636-659.

Is Search Memory Disrupted By Noise Stress?

"Search memory"

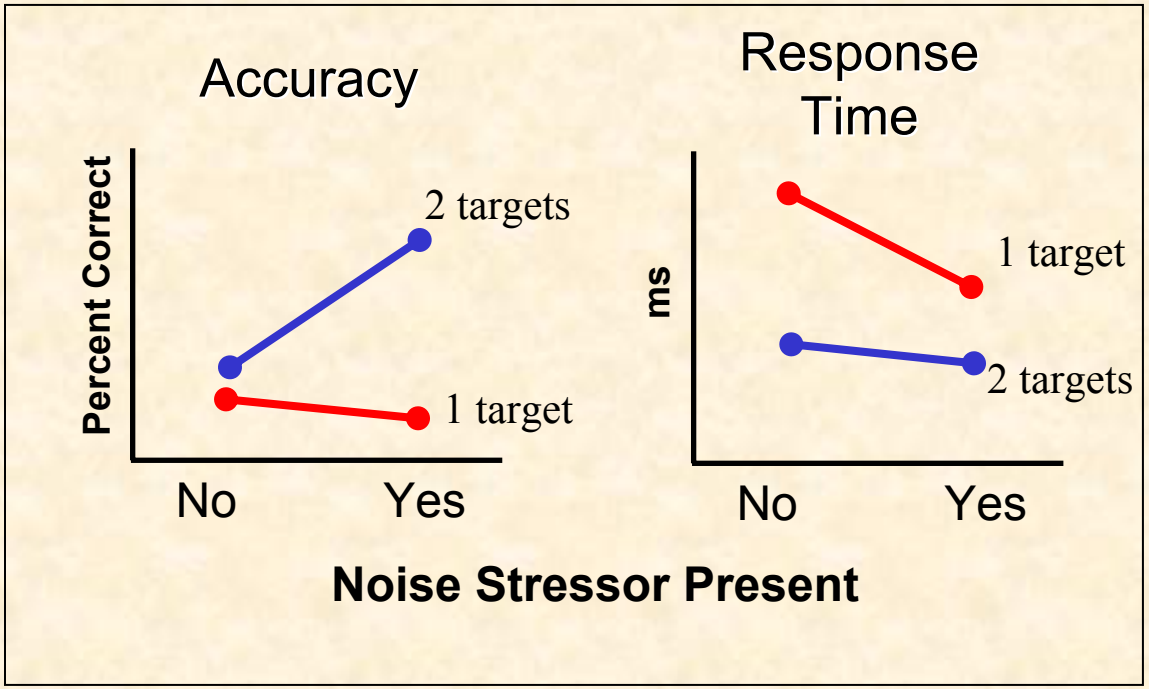
- Memory for items that have already been searched
- Facilitates efficient search
- Guides visual attention to new objects



D.D. Diaz, P.A. Hancock, & V.S. Sims. (2002). Visual Search for Two Identical Targets: The Role of Ecological Validity. Annual APA Midyear Symposium. Ft. Belvoir.

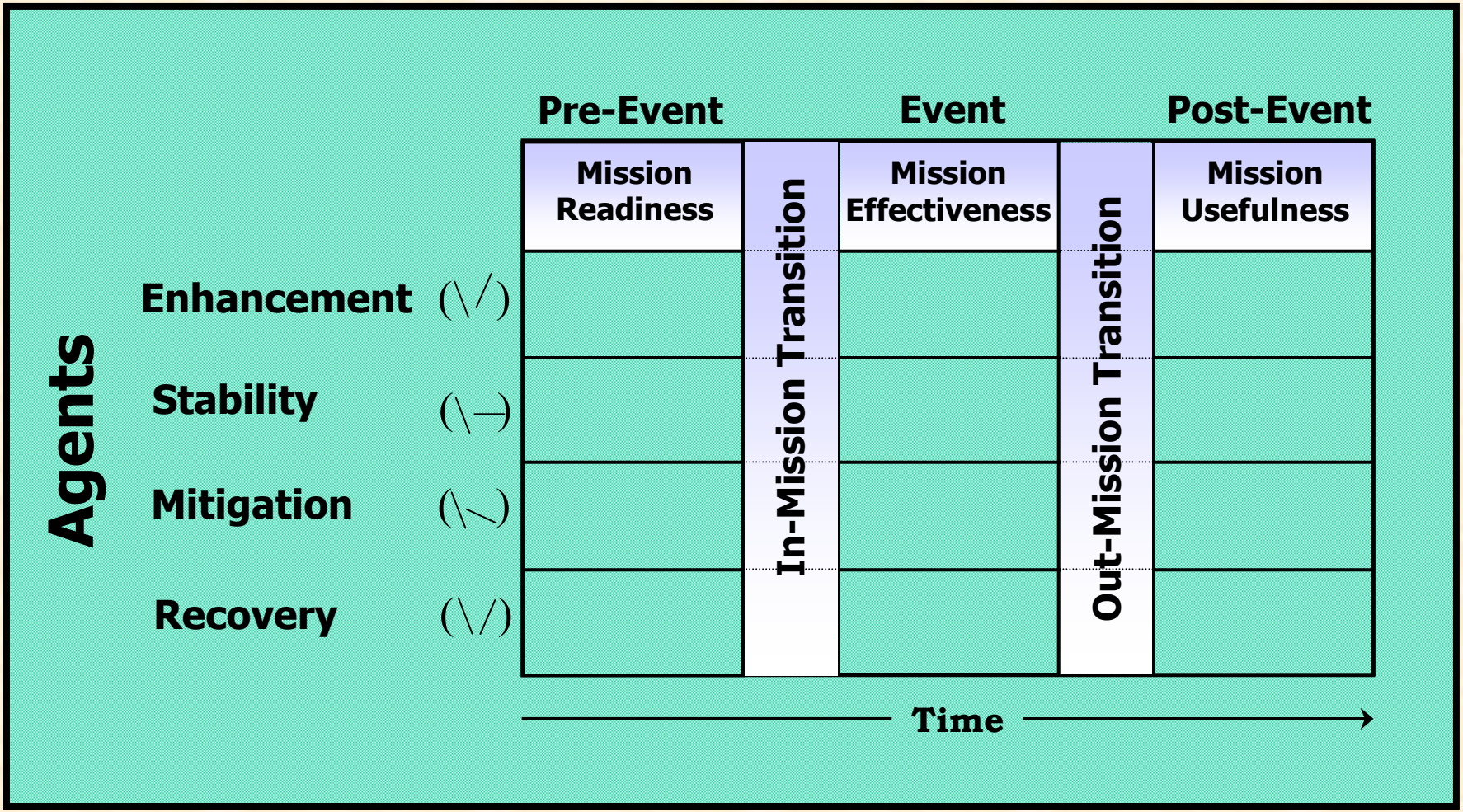
Is Search Memory Disrupted By Noise Stress?

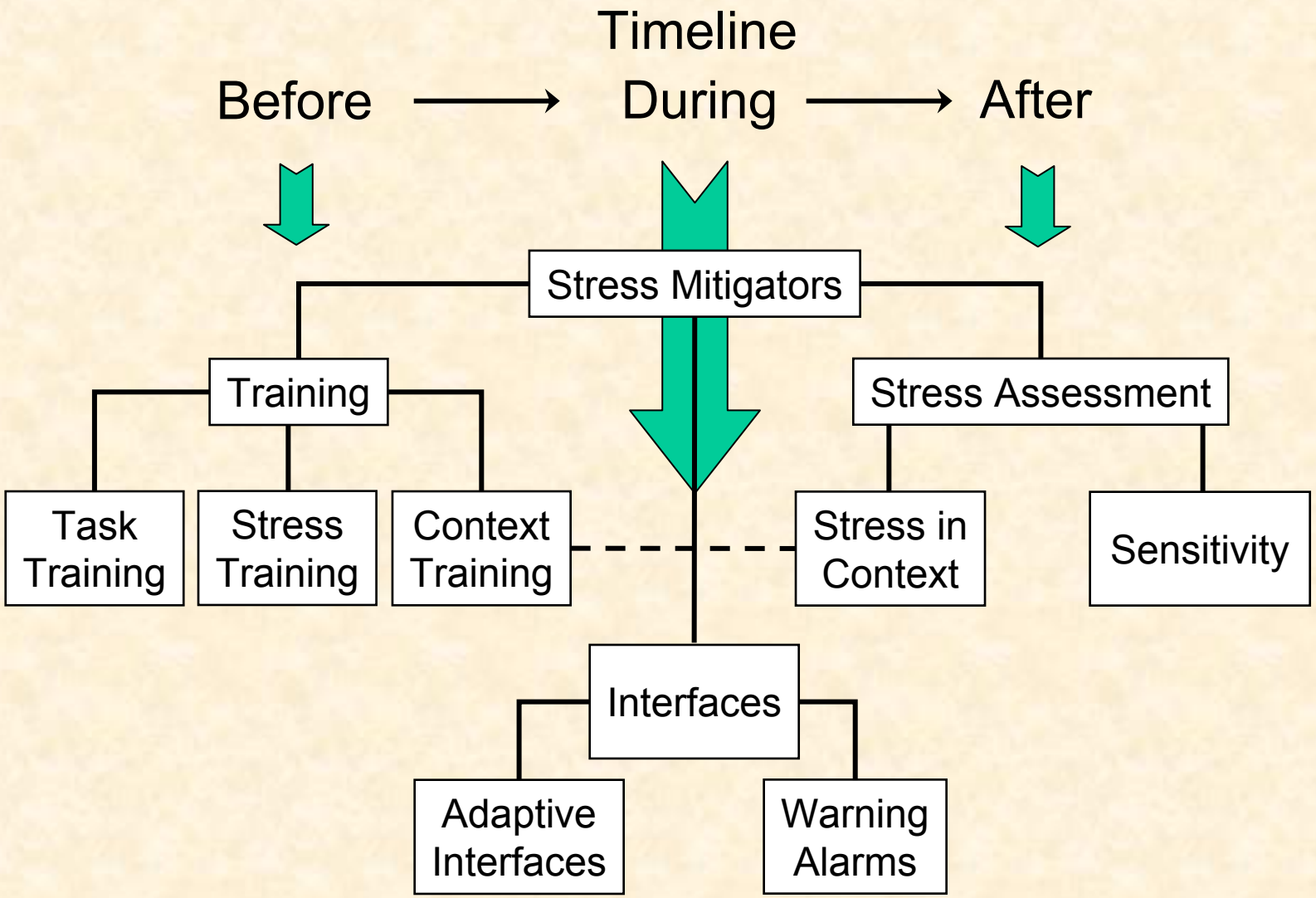




D. D. Diaz. (2002). Visual Search for Two Identical Targets: The Role of Ecological Validity. Annual APA Midyear Symposium. Ft. Belvoir.

Matrix for Performance Change




Timeline
Before
During
After
Stress Mitigators
Training
Stress Assessment
Task Training
Stress Training
Context Training
Stress in Context
Sensitivity
Interfaces
Adaptive Interfaces
Warning Alarms

OPUS Goals

- A fully articulated theory of stress and performance.
- Integration of progress into predictive models.
- Strategies for performance maintenance/enhancement under stress.
- Training approaches for such strategies.
- Interfaces predicated upon such strategic enhancements.

WEBSITE

www.mit.ucf.edu

Operator Performance Under Stress

MURI

MIT²

The Catholic University of America

Institute for Simulation & Training



Peter Hancock, Ph.D.
Department of Psychology
University of Central Florida, IST



Mustafa Mouloua, Ph.D.
Department of Psychology,



Jim Szalma, Ph.D.
Department of Psychology



Clint Bowers, Ph.D.
Department of Psychology, TPL,



Jeanne Weaver, Ph.D.
Department of Psychology



Richard Gilson, Ph.D.
Department of Psychology, CAHFA



Eduardo Salas, Ph.D.
Department of Psychology



Mark E. Koltko-Rivera, Ph.D.
Department of Psychology



Shawn Burke, Ph.D.
Department of Psychology, IST



Raja Parasuraman
Cognitive Science Laboratory
Catholic University



Arthur S. Leon, M.D.
Laboratory of Physiological Hygiene and Exercise Science
University of Minnesota.



Wayne Harris, Ph.D.
Human Factors Research Laboratory
University of Minnesota



Kip Smith, Ph.D.
Department of Psychology Kansas State University



Ava Walker, Ph.D.
Department of Kinesiology
University of Minnesota

Operator Performance Under Stress

MURI

MIT² | The Catholic University of America | Institute for Simulation & Training

UCF Human Factors Psychology Graduate Students


Christina Morris


Adams Greenwood-Ericksen


Krystal Bradley – Radakovich


H.C.Neil Ganey


Jennifer Thropp


Derek D. Diaz


Jessica Helmick


Chris Brill


Lorrenzo Torrez

Undergraduate & Post Bac. Research Assistants


Jennifer Ross


Danielle Adzima


Shawn Stafford


Chrissy Giddens


Craig Russell


Cristina Vega

Coordinator Administrative Services


Barbara Koehler
Coordinator, Administrative Services, UCF

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was Supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Operator Performance Under Stress

MURI

MIT²

The Catholic University Of America

Institute for Simulation & Training

University of Central Florida	Catholic University	University of Minnesota	Kansas State University
MIT²	**IST**		
Time Distortion	Teams Under Stress	Brain States & Stress	Physical & Mental Workload	Decision-Making
Interfaces for Stress	Stress & Team Cognition	Cognitive Workload	Physical Evaluation	Distrust
Visual Search	Low Cost Training	Viking	Viking	Biomedicine
Delta Force	Flashpoint			
Viking				

MURI Affiliations

Dr. Rene de Pontbriand,	ARL, Aberdeen Proving Ground,	Theory and Evaluation
Dr. Laurel Allender,	ARL, Aberdeen Proving Ground,	Modeling and Prediction
Dr. Linda Fatkin	ARL, Aberdeen Proving Ground,	Physical/Cognitive Workload
Dr. Jerry Krueger	The Wexford Group	Soldier Performance
Dr. John Deaton	CHI Systems	Adaptive Interfaces
Dr. Robert Hockey	University of Leeds	Theory and Modeling
Dr. Sherry Tove	ARO	Program Development
Dr. Mike Drillings	ARI	Program Advisement
Dr. Elmar Schmeisser	ARO	Program Direction
Col. Larry Shattuck	West Point	Advisory Council
Col. Bob Tyler	USMC retd	Advisory Council
Col. Bill Blackwood	US Army retd	Advisory Council
Dr. John Ruffner	DCS Inc	Industry Affiliate

Adaptive Experts and Automaticity: Explanations of Failing Under Pressure

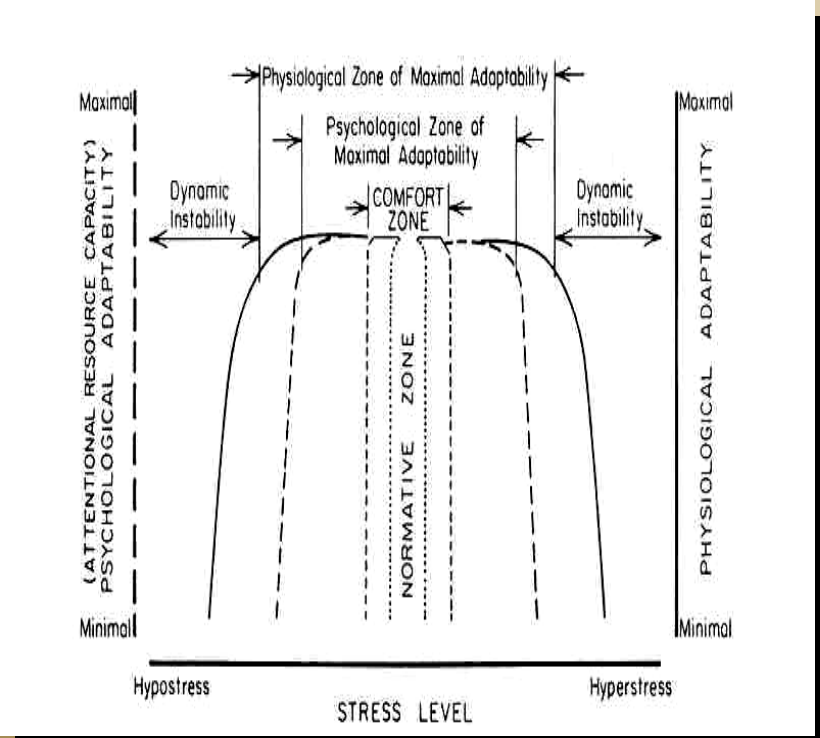
Paul Ward

James L. Szalma

Peter A. Hancock

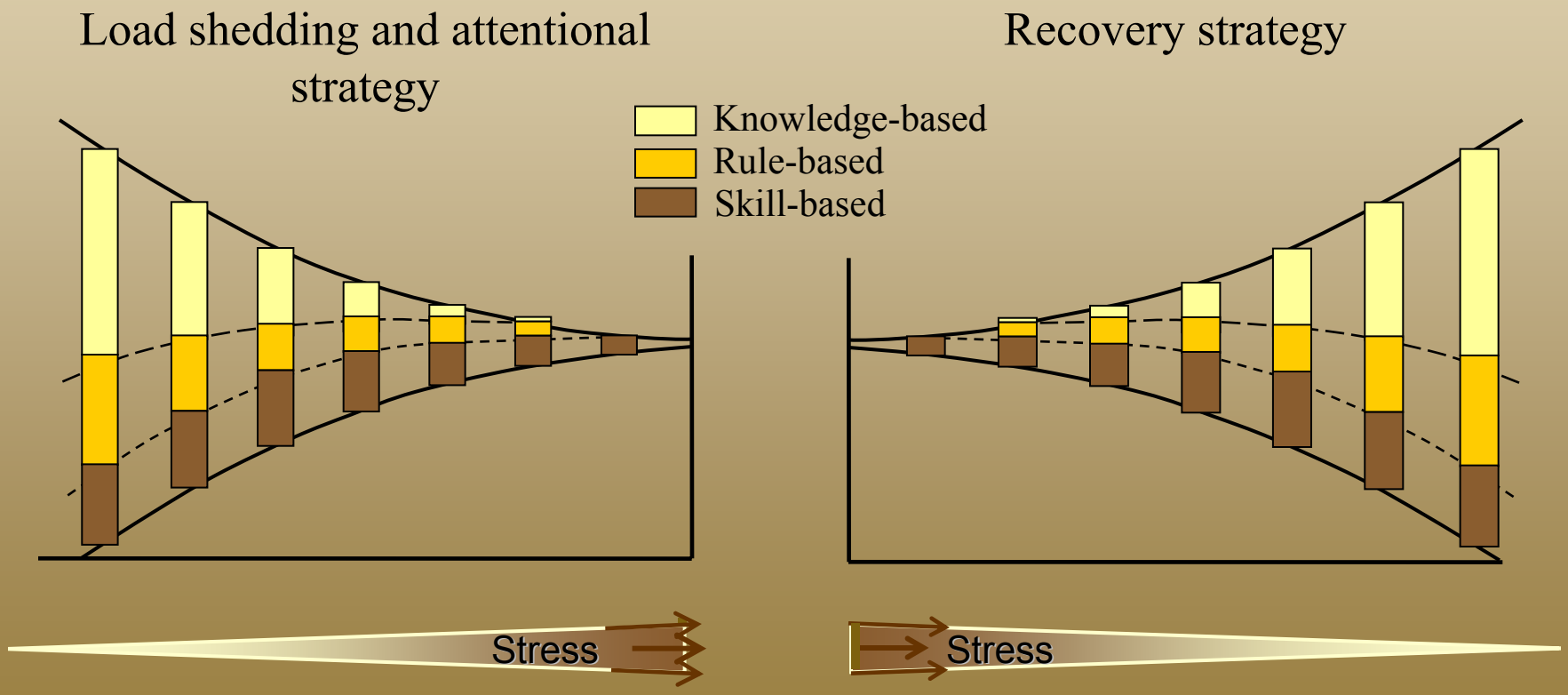
Stress and Task Performance

- Increased stress reduces the available strategies and number of solution paths to meet the goal
- Stereotypical behavior increases as attention narrows to find most appropriate and available task solution
- Individuals use compensatory strategies to maintain optimum information flow
  - e.g., modification of information structure and rate



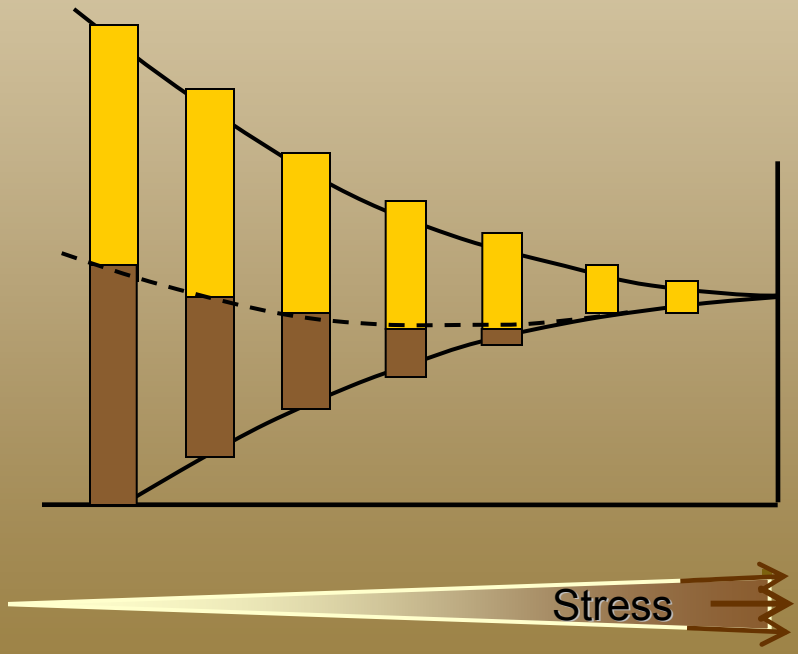
(Hancock & Warm, 1989)

Predicted Effects of Stress on Task-type/Skill-level



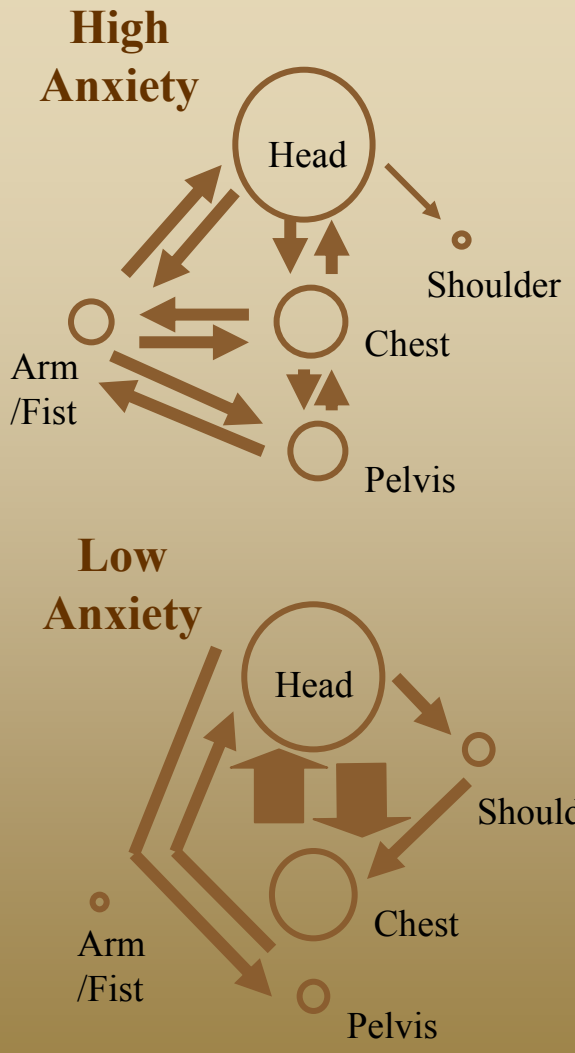
Do ‘Skills’ Break-down Under Pressure?

- Choking and conscious control (Baumeister, 1984; Masters, 1992; Beilock & Carr, 2001)



  - Lapse into conscious monitoring and control of previously automated skills under stress
  - Attention to step-by step component processes of a proceduralized skill may inhibit automated processes





Distraction Effect on Visual Search in Combat & Shooting Environments

Williams & Elliott (1999). Anxiety, expertise, and visual search strategy. JSEP, 21, 362-375.

Janelle et al. (2000). Expertise differences in cortical activation and gaze behavior during rifle shooting. JSEP, 22, 167-182.

Distraction & Conscious Control

- Hypervigilance and distractability effects in high trait anxious individuals (Eysenck, 1992)
- Attentional, Interpretive and explicit memory biases in high trait anxious individuals (Eysenck et al., 1987; Matthews & MacLeod, 1985)
- High trait anxious may invest more 'on-task' effort (Eysenck & Calvo, 1992)
- Self-focus hypothesis may be more credible than distraction hypothesis as a stress mechanism (Lewis & Linder, 1997)
  - Attending to a distracter ameliorated performance degradation under stress

➢ Increased self-awareness of explicit control (Baumeister, 1984), distraction from, and elimination of explicit control mechanisms (via implicit learning) (Masters, 1992) have similar effect in ameliorating stress.

Adding ‘Skill’ to the Equation

- Divergent predictions for distraction and conscious control dependent upon stage of learning (Beilock & Carr, 2001)
  - Distraction by task-irrelevant stimuli may disrupt performance early in learning
  - Conscious control/explicit monitoring may inhibit automated processes later in skill development

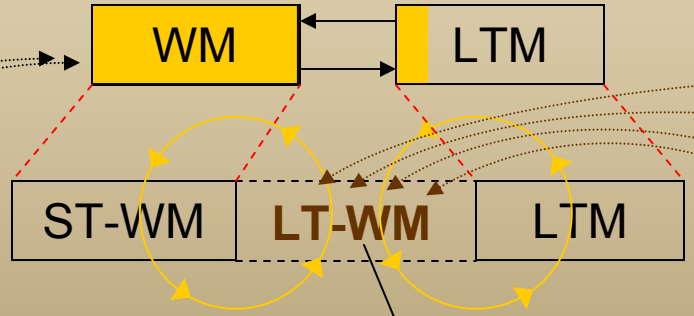
- Self-regulatory control system based in WM (Hockey, 1986)
  - Self regulatory mechanisms detract from learning in novices but aid performance in skilled individuals (Kanfer & Ackerman, 1989)

Simple Assumptions Underlying Acquisition of Expertise

1. Experts become more automated (?)
   i. Progress through cognitive to associative to autonomous stages of learning (Fitts & Posner, 1967)
   ii. Shift from controlled to automatic processing (Shiffrin & Schneider, 1977).
2. Decrease in attentional demand as a function of skill (?)

- Departure from standard view of automatization
  - Structural change through expertise development where new representations are acquired for monitoring and control (Ericsson & Lehmann, 1996)
  - Additional controlled processing with skill incl. more strategic activities, and monitoring (and adaptation) of automated processes (Shebilske et al., 1999).
- Exclusion of emotional mechanisms of stress and performance

Expertise Adaptations



Expert adaptation to WM

Storage in LTM & circumvent WM limits

Retrieval structure & elaborative encodings

Explicit monitoring, control, planning, prediction, anticipation, evaluation in skilled processes/performance

Resource-limited CE (decision making, planning, etc.) & transient, storage in WM sub-systems

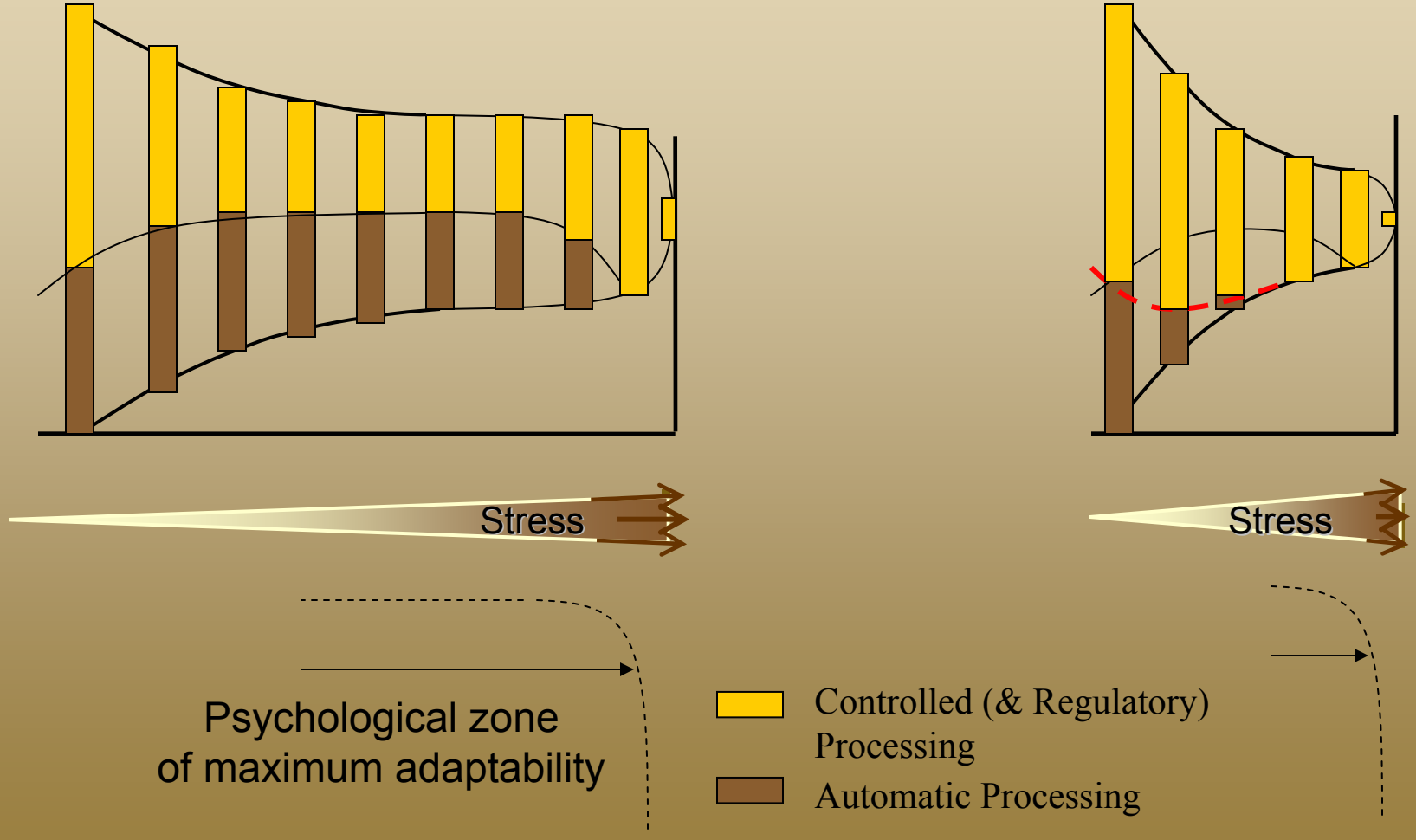
Self regulatory control system (assessment, evaluation, resource re-allocation)

Integrated cognitive and regulatory structure/mechanism

Questions for further research

- Self-regulatory control & skill-based mechanisms
  - ➢ Are they:
  - – Parallel and independent?
  - – Parallel and resource inter-dependent?
  - – Integrated mechanism?
    - Expertise mechanisms are consistent with an integrative mechanisms view (Ericsson & Lehmann, 1996; Ericsson, 1998; Shebilske et al., 1999)
      - – No explicit test of this hypothesis
  - ➢ Are the experienced levels of stress dependent upon qualitative differences in the nature of the domain-specific skill mechanism involved, type of encoding/processing, or the nature of the skill itself?
  - ➢ To what extent are the acquisition of integrative mechanisms dependent upon individual differences?

Expert-Novice Differences in Response to Stress



Fuzzy Signal Detection Theory: Initial Evidence for a Novel Approach to Evaluation of Decision-Making Under Uncertainty

J.L. Szalma and P.A. Hancock
University of Central Florida

For almost fifty years Signal Detection Theory (SDT) has provided an extremely useful analytical tool for evaluating human and machine performance in both simple and complex systems. However, in the SDT model the state of the world is organized into crisp mutually exclusive categories (i.e., signal versus non-signal; old versus new; etc.), which may not be accurate representations of the true states of the world that face operators who must detect critical events. To capture the uncertainty inherent in real-world situations the representation of the stimulus dimension must be permitted to be continuous. A recent advance in SDT promises to enhance measurement of performance in these domains by capturing the continuous-nature of real-world stimuli and also by incorporating into the model the uncertainty inherent in such stimuli and in the operators who respond to them. This development, Fuzzy Signal Detection Theory (FSDT; Parasuraman, Masalonis, & Hancock, 2000), combines traditional SDT with Fuzzy Set Theory to extend signal detection analysis beyond the traditional decision-making model. FSDT permits events to simultaneously be in more than one state (e.g. signal and non-signal), so that a particular instantiation of a stimulus (e.g. randomly drawn from a stimulus continuum) has properties of both 'signal' and 'non-signal.' This allows stimulus and response dimensions to be continuous rather than categorical. Consequently, FSDT can be employed in applied settings where the degree to which an event is a signal for detection may vary and decisions must be made under conditions of high uncertainty. Recent evidence indicates that FSDT provides a better description of the effects of stimulus uncertainty on observers' responses, while the traditional model is more sensitive to manipulation of task difficulty (Murphy, Szalma, Hancock, & Stafford, 2003). Further work is necessary to investigate the relation between task difficulty and the FSDT model. In addition, there is evidence that the FSDT model captures the shifts in criterion setting resulting from changes in the distribution of 'signal-like' stimuli (Szalma, Stafford, & Hancock, 2003), similar to patterns observed in prior vigilance research in which changes in signal probability were evaluated using the traditional SDT model (e.g. see Parsons et al., 2000; Vickers & Leary, 1983). Future research will further examine the FSDT model by investigating the performance effects of variations in the distribution of stimulus uncertainty.

Abstract presented at the ARO Bioscience Workshop, "Seeing Both the Scene and Unseen-Human Sight and Microscopic Sites," in Cashiers, NC April 2003.

References

Murphy, L.L., Szalma, J.L., Hancock, P.A., & Stafford, S.C. (2003, March). *Comparison of methods of analyses from fuzzy and traditional signal detection theory*. Paper presented at the Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society, Ft. Belvoir, VA.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors,* **42**, 636-659.

Parsons, K.S., Warm, J.S., Matthews, G., Dember, W.N., Galinsky, T.L., & Hitchcock, E.M. (2000, November). Changes in signal probability and vigilance performance. Poster presented at the meeting of the Psychonomic Society, New Orleans, LA.

Szalma, J.L., Stafford, S.C., & Hancock, P.A. (2003). Application of fuzzy signal detection theory to vigilance: The effect of criterion shifts. Manuscript under review.

Vickers, D., & Leary, J.N. (1983). Criterion control in signal detection. *Human Factors*, **25**, 283-296.

Abstract presented at the ARO Bioscience Workshop, "Seeing Both the Scene and Unseen-Human Sight and Microscopic Sites," in Cashiers, NC April 2003.

DOD Multidisciplinary Research Program: MURI Operator Performance Under Stress (OPUS).

White Paper

Stress, Human Information Processing, and Performance Mediators: Application of a Descriptive Framework to Current Modeling Tools

P. Ward, J.L. Szalma, and P.A. Hancock.

Department of Psychology and Institute for Simulation and Training

University of Central Florida

2nd Draft: 11-25-02

Work performed under the U.S. Army Research Office – MURI: Optimizing Cognitive Readiness Under Combat Conditions

Version 2.1

Stress, Human Information Processing, and Performance Mediators: Application of a Descriptive Framework to Current Modeling Tools

Abstract

A three dimensional matrix is proposed which extensively details the interactive effects of stress, information processing and moderators of the stress-performance relationship. The aim of this descriptive framework is to serve as a meta-analytic review of the effects of stress on human performance and to inform current modeling practice used to predict soldier-system performance under stress. The U.S. ARL-HRED presently employ a task network modeling architecture (IMPRINT) in this regard which is largely used for procurement purposes in determining the requirements of human performance in real-world contexts. Within IMPRINT, performance on military occupational specialty tasks is currently modeled under standardized conditions using the ACT-R cognitive architecture. Performance shaping functions and stress degradation algorithms are subsequently applied to the modeled data to reflect individual differences in task performance under stressful conditions. Whilst IMPRINT has been successfully implemented, it is somewhat limited both by the availability of suitable stress degradation algorithms and the broad brush approach used to apply the same function/algorithm to all tasks grouped together by a taxonomic classification system. The matrix aims to provide resolution, at least in part, to a number of issues in this regard. Firstly, by clearly outlining the multitude of environmental, physical and task-related stressors within the matrix, and considering the transactional relationship between operator and task, a more complete understanding of stress effects on human performance can be delineated. Similarly, a comprehensive description of the effects of stressors on a range of information processing variables sub-serving, perception, cognition, and action will allow both theoretical and practical implementation issues related to current modeling approaches to be addressed. In addition, the inclusion of moderator variables within the matrix will provide an added dimension that will facilitate future modeling endeavors.

Introduction

A plethora of research exists which documents the effects of stress on human performance (see Hancock & Desmond, 2001). However, much of this research has focused upon the effects of a single stressor on one specific micro-cognitive processes (e.g., selective attention, working memory). To date, very few assessments have been made on a broad range of interactive stressors with the specific intention of detailing both the variable and predictable effects upon perception, cognition, and action. Hockey and Hamilton's (1983) efforts to discern the difference between the cognitive patterning of stress states represents one exception in this regard. These authors considered the effects of multiple sources of stress (e.g., anxiety, noise, narcotic, etc.) on strategic (i.e., attentional selectivity, decision speed/accuracy) and structural (i.e., short-term memory, alertness) information processes. Their findings indicate that although some commonality exists between particular stressors, a multidimensional approach was necessary to explain the lack of response uniformity. While some of the conclusions made by Hockey and Hamilton were 'reliable and substantiated' the authors acknowledged that on occasion, interpretation was based upon 'rather isolated results' (p.350). The results were also not discriminated by the magnitude of effect further decreasing the clarity of interpretation. Notwithstanding, Hockey and Hamilton's research provides a first step to making a concerted effort in determining a range of stressor effects on various information processes within a single investigation.

Lazarus and Folkman (1984) proposed that stress relates to the person-environment interaction. This transactional viewpoint suggests that it is an individual's appraisal of environment as taxing or exceeding available resources that moderates the effect of a stressor. Matthews (2001) recently suggested that this transaction occurred at multiple levels (e.g., biological, psychological). The negative effects of stress occur when individuals appraise an event as threatening or endangering their well being and when they consider that available coping strategies would be ineffective in dealing with the stressor. Consequently, those variables that moderate the relationship between sources of stress and the nature of task-related information processing are likely to be influenced by the transaction between operator and task. Although a wealth of literature exists on coping (e.g., Matthews & Campbell, 1998; Smith & Lazarus, 1993) and effort allocation (e.g., Eysenck & Calvo, 1992), surprisingly, very little empirical research has specifically attempted to identify those variables that moderate stress effects on task-related information processing in a real world setting. Bowers, Weaver, and Morgan (1996) examined the stress-performance relationship and highlighted a number of potential variables that may act as stress moderators. These included social support, locus of control, perceived control, trait anxiety, self-efficacy, self control and experience. However, further research and a more encompassing approach to consolidating the known effects of moderator variables is needed in this area.

To fill the current void, we propose a descriptive framework which details the effects of a broad range of stressors on information processing and examines the potential variables that may moderate this interaction. Our aim is to provide both a descriptive database and meta-analytic review that informs current mechanisms for predicting task

performance under stress. Ultimately, the purpose of this venture is to increase the current prediction capacity of existing models of human performance that incorporate some form of stress degradation algorithm, moderation process or workload assessment. To meet these goals, a three-dimensional matrix has been constructed which encapsulates the variables of interest on each dimension: sources of stress, information processes, and moderators of the stress-performance relationship. Before giving significant detail on each dimension, an overview of the current models of soldier-system and human cognitive performance prediction is provided.

IMPRINT

Improved Performance Research Integration Tool (IMPRINT) is a task network modeling architecture developed by the U.S. Army Research Laboratory Human Research and Engineering Directorate (ARL-HRED). It is used to represent soldier-system performance for the purposes of test and evaluation. Two of the primary aims of IMPRINT are to determine the soldier performance requirements for the acquisition of a new system and in particular, to predict soldier-system performance under a myriad of conditions (Allender, Salvi, & Promisel, 1997; Lebeire, Biefeld, Archer, Archer, Allender, *et al.*, 2002; O'Brien, Simon, & Swaminathan, 1992). Earlier DOS-based versions of IMPRINT (e.g., HARDMAN III) were constructed from discrete modules for a range of operator-, task-, and system-related estimations. IMPRINT is now implemented using the Microsaint task network modeling environment which subdivides system missions into functions and tasks, and produces performance estimates and comparisons to a standard. Given that a multitude of performance changes are likely to occur under stressful compared to standardized conditions and that some individuals may be more suited to a particular task than others, of specific interest in these and subsequent operations are the performance shaping functions (PSFs), stress degradation algorithms (SDAs) and workload modeling tools used in performance prediction. The PSFs predict relative changes in performance for a specific set of personnel characteristics and training regimen (i.e., recency and frequency of practice) whereas the SDAs (or look up tables for stressors) modify typical/baseline measures to account for negative affects upon performance. Current methods of workload prediction will be addressed in a later section.

The relationships between predictor and performance for both PSFs and SDAs are described for categories of tasks (i.e., taxons) rather than specific tasks. That is, different functions/algorithms were developed for different taxons (see O'Brien et al., 1992). Similar effects for tasks that fall into the same taxon can be accounted for using this method. The task taxonomy was largely borrowed from Berliner, Angell, and Shearer (1964) with attempts made to take Wickens' (1980) structure for processing resources into consideration (see also Fleishman & Quantance, 1984). Nine taxons were identified spanning perceptual, cognitive, psychomotor and communication modalities:

- visual recognition/discrimination
- numerical analysis
- information processing/problem solving
- fine motor (discrete)

- fine motor (continuous)
- gross motor (heavy)
- gross motor (light)
- communication (reading and writing)
- communication (oral)

Tasks relevant to each Military Occupation Specialty (MOS) are user-assigned to a maximum of three taxons. Personnel characteristics such as the Armed Services Vocational Aptitude Battery (ASVAB) composite scores, reading grade level, complex perceptual speed/accuracy, simple reaction speed/accuracy, etc. are used to predict task performance for all tasks that are pre-assigned to each taxon by means of the PSFs (which translate to a series of regression equations based on aggregated and standardized test scores). Due to the differential effects of stressors on each task, tasks are weighted according to the taxon to which they have been assigned so that, when the SDA is applied (post application of the PSF), the impact of stressors on each task is accordingly weighted (for a more detailed overview of SDAs and PSFs, see Allender et al., 1997; O'Brien et al., 1992).

At present, five, predominantly environmental, stressor equations exist within IMPRINT. These relate to Mission Oriented Protective Posture (MOPP) gear for biological, chemical and nuclear defense (i.e., protective clothing), heat, cold, noise, and sustained operation (i.e., sleepless hours). Each of these algorithms are applied individually or in combination via the use of a power function (see Harris, 1985). The majority of degradation factors for each stressor were derived from research prior to 1988, some of which dates back to the 1950's. Moreover, the SDAs that have been developed are only capable of modifying tasks within a taxon for a very limited number of stressors (i.e., one to three). In addition, where degradation algorithms have been written, a comprehensive evaluation of current stressors has not followed. Allender et al. (1997) indicated that this is a high priority for research. Even a cursory glance at the literature suggests that there is much scope for updating SDAs to incorporate 'state of the art' knowledge and consequently bridge the gap between existing stressors and their effects on current tasks/taxons. Equally important is that this information will increase the number of stressors and the associated degradation factor that could be applied to each taxon. The U.S. ARL-HRED and similarly, the U.S. Defense Special Weapons Agency are currently conducting research in this area to develop a prioritized list stressors and their effects based upon 'user need' and the 'availability of generalizable data' (Allender et al., 1997; see also Anno, Dore & Roth, 1996). Finally, the cognitive mechanisms that underlie stress effects need to be explored. For performance under standard conditions, the IMPRINT model utilizes the ACT-R theory of cognition to model performance. It is to this cognitive architecture that we now turn.

ACT-R (/PM)

Adaptive Control of Thought-Rational (ACT-R) (Anderson & Lebiere, 1998) is both a theory of cognition and a hybrid cognitive architecture that models and predicts human cognition. It operates at both symbolic (e.g., as a production system) and sub-

symbolic (e.g., as a neural-like activation-based system) levels. At the symbolic level ACT-R assumes two types of knowledge. Declarative knowledge is that which we are consciously aware of and is verbally describable, whereas procedural knowledge specifies how declarative knowledge is instantiated to solve a problem. Its basic cognitive operation is a production-rule, often referred to as a condition-action pair or 'if… then…do…' statement. ACT-R is both a parallel-matching and serial firing production system centered on the current goal or focus of attention. When the left hand side of the production (i.e., the condition) is matched with the current goal, post conflict resolution (i.e., selection of a single production from multiple productions that match information currently held in working memory), the production-rule fires and the right hand side of the production (i.e., the action) modifies the current, or deposits a new, goal on to the 'goal stack' (i.e., in working memory). As an alternative to a cognitive action, a physical action may also be executed.

Goals are represented as chunks (i.e., encoded structures with a number of slots or attributes ($\leq$ 6), each of which hold a single piece of information) and stored in declarative memory along with other information also represented as chunks. New productions are created from declarative chunks via a process of production compilation. That is, a new generalizable production cycle is generated from the residual execution trace of multiple production firings (Lebiere et al., 2002). At the sub-symbolic level, retrieval of chunks from declarative memory is based upon the chunk with the highest level of activation. Activation within this framework is derived from the attentional weight given to the goal based upon factors such as history of use, information contained within a chunk, and time-based decay. A set of sub-symbolic learning processes can also modify activation weights. In a similar manner, conflict resolution is determined as a rational process of both the estimate of a production's worth and the trade-off between the probability and cost (e.g., time) of goal achievement (for more details, see Anderson & Lebiere, 1998).

In an attempt to extend ACT-R's application to tasks that involve both perception and action as well as cognition, Byrne and Anderson (1998) first proposed ACT-R/PM as a mechanism to augment central cognition. The perceptual-motor (PM) component is now comprised of visual, motor, speech, and auditory modules (see also Anderson, Matessa, & Lebiere, 1997; Byrne, 2001) and was largely influenced by the EPIC architecture (see Kieras & Meyer, 1997), particularly at the motor level. ACT –R/PM was developed to model both the interaction with environmental stimuli and the perceptual-motor demands of most real world tasks.

ACT-R and its integration with IMPRINT

ACT-R and IMPRINT have recently been integrated for the purposes of modeling and predicting task performance in situated events. The most developed application to date has been in a pilot navigation context for taxiing along the runway (Lebiere et al., 2002). A similar integration of ACT-R and an IMPRINT-related task network modeling environment (TNME) has also recently been implemented in a shoot-list management task (e.g., recognition and decision making) by the U.S. Air Force Research Laboratory's

Human Performance Model Integration (HMPI) program using the Combat Automations Requirements Testbed (CART) TNME (see Craig et al., 2002). In both situations, ACT-R and IMPRINT functionally combine by focusing upon different aspects of performance. Whilst IMPRINT is predominantly concerned with the macro-components of real world tasks, ACT-R contends with a task's cognitive, perceptual, and motor components.

To direct ACT-R's goal driven processes, task-related information is sent from IMPRINT to ACT-R via a component object model (COM) link which specifies a goal in ACT-R. This in turn creates a goal set that corresponds to the task description. Each sub-goal specified within ACT-R results in data regarding how the goal was accomplished, as well as the latency and accuracy (i.e., number of errors) of each goal's completion. Latencies for retrieval are determined as a function of a chunk's activation whereas the time taken to apply a production is determined by a production's action component. Rudimentary internal actions take approximately 50 ms, while external actions typically take longer and are determined by the ACT-R/PM module.

Errors occur due to the stochastic nature of the decision process and its interaction with a dynamic environment. As opposed to errors being a problem of 'faulty symbolic logic' (Lebiere et al., 2002, p.3), the inherent variability is more likely to reflect an accurate representation of human operators' performance under changeable conditions than that of a purely deterministic system. Errors may originate from a number of sources within ACT-R such as retrieving the wrong information or failure to retrieve necessary information from memory (i.e., memory failures), use of an ineffective production rule (i.e., choice or selection errors), or failure of the perceptual modules to detect important information (i.e., attentional errors).

One of the benefits of using this type of integration is the reduced need for endless data. Moreover, ACT-R can generate distributions of time and error data for each task within IMPRINT that appropriately reflect the actual perceptual, cognitive, and motor components of tasks experienced in the real world, albeit under standardized conditions (for detailed summary, see Lebiere et al., 2002).

In addition to modeling task performance, associated workload is currently being modeled within ACT-R (see Lebiere, 2001; Lebiere et al., 2001) such that the level of effort required to perform each MOS task could be used to moderate the 'attributes of the memory model' (Lebiere et al., 2002, p.6). As an adjunct to the IMPRINT/ACT-R integration model, recent developments have been made in modeling related factors such as emotion (e.g., Belavkin, 2001), behavior moderators (e.g., Ritter et al., 2002), and fatigue (e.g., Jongman & Taatgen, 1999) within ACT-R. There is no current initiative to replace IMPRINT's performance shaping functions and stress degradation algorithms with an ACT-R (/PM) model. Regardless, the potential exists to model the effects of stress in this manner. Accordingly, when determining the effects of stressors and moderators on perceptual, cognitive and motor mechanisms current and potential modeling developments need to be considered.

The Matrix

'The matrix' is a three-dimensional entity which has been constructed to inform the functional content, and where necessary, modify the structural basis of current modeling applications used in soldier-system performance prediction. The impetus for this research and review process is to increase current prediction capacity under stressful conditions by informing the construction of performance shaping functions and degradation algorithms for existing and future stressors within IMPRINT. In addition, to ensure that existing research is compatible with the adopted theoretical framework and modeled data derived via ACT-R.

The first dimension of the matrix consists of a complete catalog of performance stressors. The second dimension details the range of human information processes that could potentially be affected by these stressors. Moderators of the stress-performance relationship reside on the third dimension. Collectively, the matrix provides an integrative database that will drive theoretical development, facilitate future modeling efforts and allow existing models of stress and performance (e.g., Hancock & Warm, 1989) to be tested.

The stress dimension

As documented earlier in this paper, IMPRINT currently utilizes data for five stressors, and in the worst case scenario, employs only one stress degradation algorithm for a task that falls within a particular taxon. Clearly, the current implementation is not exhaustive and other environmental (e.g., social), physical (e.g., fatigue, hunger) and task-based stressors need to be incorporated within this dimension to make it comprehensive. Hancock and Warm (1989) suggested that tasks themselves are likely to be a source of stress, particularly in terms of information overload and as such should be appropriately considered within this dimension. One of the goals of this research was to build upon the seminal work of Hockey and Hamilton (1983) to determine both the dynamic and stochastic nature of the cognitive patterning of some stressors while potentially identifying the deterministic, or at least invariant, performance effects for others. Hockey and Hamilton's approach fits well within the IMPRINT model in that it mirrors the taxon system (e.g., particular stress effects for certain types of tasks) devised within IMPRINT to which stress degradation algorithms are applied. However, given the current delineation of taxons, for instance, including all numerical analysis, information processing and problem solving tasks in to one taxon, it is likely to prove impossible to derive a 'common' stress degradation algorithm for each taxon within the existing categorization scheme. Moreover, there is concern regarding the reliability of user assignment of tasks to taxons. Christensen and Mills (1967) highlighted a median inter-rater reliability of .78 (Rho range = .29 to 1.00). This has enormous implications for applying SDA's to tasks, where differences between users in task assignment could lead to different SDA's being applied and contradictory predictions on performance for the same task.

The information processing dimension

Information processes from perception to cognition to action are mapped onto this dimension. This is probably the most complex of dimensions for a number of reasons. At

a gross level of cognition, most cognitive and human factors psychologists and cognitive scientists would agree on the existence of general processes such as problem solving, reasoning, judgment and decision making. However, with a more fine-grained analysis of cognition comes a vast expanse of theoretical conflict in terms of viewpoints, terminology, systems, mechanisms and processes. On a practical implementation level, those perceptual and cognitive processes which are relatively generic such as stereoscopic depth sensitivity, encoding, recall and recognition sit comfortably along this dimension without much discord. In contrast, for those processes that are more theoretically bound it is necessary to create multiple columns and to index appropriate research under each relevant column. As an example, research that has examined the effects of stress on working memory could arguably be placed into columns which correspond to both short term and working memory.

On a theoretical level, very little 'human' research has been conducted on the effects of stressors on procedural and declarative memory or on any of the sub-processes defined within the ACT framework. Even less research exists that has successfully modeled human stress and emotion in ACT-R (for exceptions, see Belavkin, 2001). Considering the questions raised by Szalma and Hancock (2002) regarding the effects of stress on Rasmussen's (1983) trichotomy of skill, rules, and knowledge, the application of stress degradation algorithms post modeling of performance via both procedural and declarative mechanisms may be a contentious issue. To accurately model the actual stress effects upon human information processing, algorithms may have to be included as part of the cognitive modeling process, not attached *a posteriori*. Moreover, without empirical investigation, the level of parity is unclear between research examining the effects of *stressors* on, for instance, tasks involving short-term memory as proposed by Atkinson and Shiffrin (1968) or those involving working memory as proposed by Baddeley (1986) and the application of stressor degradation algorithms on task data that is derived via ACT-R computation. Consequently, it is an open question whether the resultant effects of stressors are directly comparable given that underlying mechanisms differ both structurally and functionally at a micro-level of analysis. While there are clear, but often subtle, conceptual differences between many overlapping constructs, practically speaking, this argument may turn out to be a straw man. However, empirical research is required to test the relevance and validity of existing stress-related phenomena to current methods. The proposed matrix will guide such research.

The third issue is an extension of the problem posed under the stressor dimension. That is, other than at the perceptual, cognitive, and psychomotor level only modest parity exists between taxons and the vast array of information processes upon which research has been conducted. Using the current taxonomy as a template for dimension division is likely to result in a disproportionate and asymmetrical structure that conceptualizes the majority of processes as essentially equivalent and consequently, has little explanatory power. Although the original study suggested that the task taxonomy was developed with Wickens' (1980) processing structure in mind (see O'Brien et al., 1992), there is little which resembles this model and that distinguishes it from Berliner et al.'s (1964) original classification scheme. Again, modifications to the current taxonomy may be necessary to resolve this issue.

The performance moderator dimension

Unfortunately, the least amount of research has been conducted in this area. As was suggested in the introduction, the person-environment interaction is likely to play an active role in moderating the effects of stress between operator and task. Consequently, coping and appraisal are likely to make up a significant proportion of this dimension. However, related factors, for instance, perceived control (see Bowers et al., 1996), resource allocation policy in terms of compensatory effort (e.g., Kahneman, 1973; Matthews & Campbell, 1998; Sanders, 1983), training (e.g., Hall-Johnston & Cannon-Bowers, 1996), personnel selection (e.g., Hogan & Lesser, 1996), and system design itself (e.g., Wickens, 1996) may also moderate the effects of this interaction. Within the context of IMPRINT and ACT-R, personnel characteristics and training data used in the performance shaping function, and workload estimates are all currently used to moderate the stress-performance relationship.

Concluding Remarks

The construction of the matrix is as labor intensive as any meta-analytic review process. However, the purpose of its development is not for meta-analytic purposes alone. The primary purpose of its development is to improve the integrated model's capacity to predict soldier-system performance. The matrix will provide much needed information on the effects of stressors on information processing and those variables that can be used to moderate the process. Empirical verification of modeled cognitive processes will facilitate integration of stress-related effects into future modeling ventures. The integration of new stressors, validation and modification of old stressor effects, and revisions to the existing taxon scheme, once integrated, are hypothesized to all provide an immediate effect upon performance prediction. However, data can only inform where there is data to inform. The glaring truth of the matrix's construction is likely to be that there are as many gaps in the current knowledge base as there is understanding of these complex effects. These gaps need to be filled through further empirical research. Where knowledge is lacking however, system design can still be enhanced by the matrix. Where data does not exist for the effects of particular stressors on certain processes then design emphasis can be placed on taxing those systems where the stress effects are known.

Identified Research Questions

- Can existing models of stress and human performance account for effects of stress identified by the matrix?
- How well do current stress degradation algorithms used within the IMPRINT model predict real-world performance under the stressful conditions to which they apply?
- To what extent can current stress degradation algorithms be modified by the matrix to predict known stress effects?
- Can stress effects on information processing be delineated in a meaningful fashion given the abundance of theoretical models of performance and cognition?

- As the information processing dimension of the matrix is specified, how will this information fit with the existing taxon classification scheme?
- How does the effect of stress on skills rules, and knowledge impact the current method of modeling cognitive, perceptual and motor performance before stress degradation algorithms are applied?
- Can a consensus of the known stress effects on information processing tasks be reached to fit within the current taxonomic classification system?
- Will the differential stress effects on information processing tasks necessitate a modification of the existing taxon scheme? If so, can a series of appropriate stress degradation algorithms be derived to fit a new scheme?
- Can the effects of stress be modeled within the IMPRINT/ACT-R integration model?
- What are the main variables moderating the effects of stress on task performance?
- Can the effects of moderating variables be accounted for in the IMPRINT/ACT-R integration model, as per the current workload implementation?
- What is the commonality of moderator effects across information processes, tasks, and taxons?
- What gaps exist in the literature regarding the stress-performance relationship and the effects of moderator variables?

Anticipated Outcomes

- A meta-analytic review of stress effects, including the moderating effects of several variables (e.g., training, operator traits, etc.) will inform and enhance modeling capacity and theory development.
- The descriptive framework can be directly applied to the mechanisms used by IMPRINT resulting in stronger prediction capacity of this model of performance under stressful conditions.
- Preliminary steps toward an integrative theoretical and modeled account of stress effects on MOS tasks/taxons.

References

Allender, L., Salvi, L., & Promisel, D. (1997). Evaluation of human performance under diverse conditions via modeling technology. In *Proceedings of Workshop on Emerging Technologies in Human Engineering Testing and Evaluation*, June 1997. Brussels, Belgium: NATO Research Study Group.

Anderson, J.R. & Lebiere (1988). *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum.

Anderson, J.R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and visual attention. *Human Computer Interaction*, 12, 439-462.

Anno, G.H., Dore, M.A., & Roth, T.J. (1996). *Taxonomic model for performance degradation in combat tasks*. (DNA-TR-95-115). Alexandria, VA: Nuclear Defense Agency.

Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system of its control processes. In K.W. Spence and J.T. Spence (Eds.), *The psychology of*

learning and motivation: Advances in research and theory, vol.2, (pp.89-187). New York: Academic Press.

Baddeley, A.D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Belavkin, R. V. (2001). The Role of Emotion in Problem Solving. In *Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing* (pp. 49-57). Heslington, York, England.

Berliner, D.C., Angell, D., & Shearer, J.W. (1964). Behaviors, measures, and instruments for performance evaluation in simulated environments. *Paper presented at a symposium and workshop on the quantification of human performance*, August. Alburquerque, New Mexico.

Bowers, C.A., Weaver, J.L., & Morgan, B.B. (1996). Moderating the performance effects of stressors. In J.E. Driskell and E. Salas (Eds.), *Stress and human performance.* (pp.163-192). Mahwah, NJ: Lawrence Erlbaum.

Byrne, M.D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies*, 55, 41-84.

Byrne, M.D. & Anderson, J.R. (1998). Perception and action. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought*, (pp. 167-200). Mahwah, NJ: Lawrence Erlbaum.

Craig, K., Doyal, J., Brett, B., Lebiere, C., Biefield, E. *et al.* (2002). Development of a hybrid model of tactical fighter pilot behavior using IMPRINT task network modeling and the Adaptive Control of Thought-Rational (ACT-R).

Eysenck, M.W. & Calvo, M.G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion*, 6, 409-434.

Fleishman, E.A. & Quaintance, M.K. (1984). *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic Press.

Hall-Johnston, J. & Cannon-Bowers, J.A. (1996). Training for stress exposure. In J.E. Driskell & E. Salas (Eds.), *Stress and human performance* (pp.223-256). Mahwah, NJ: Lawrence Erlbaum.

Hancock, P.A. & Desmond, P.A. (2001). *Stress, workload, and fatigue*. Mahwah, NJ: Lawrence Erlbaum.

Hancock, P.A. & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.

Harris, D. (1985). A degradation methodology for maintenance tasks. Alexandria, VA. HQDA, MILPERCEN (DAPC-OPA-E).

Hockey, R. & Hamilton, P. (1983). The cognitive patterning of stress states. In G.R.J. Hockey (ed.), *Stress and fatigue in human performance* (pp.331-362). New York: John Wiley & Sons Ltd.

Hogan, J. & Lesser, M. (1996). Selection of personnel for hazardous performance. In J.E. Driskell & E. Salas (Eds.), *Stress and human performance* (pp.223-256). Mahwah, NJ: Lawrence Erlbaum.

Jongman, L. & Taatgen, N. A. (1999). An ACT-R model of individual differences in changes in adaptivity due to mental fatigue (pp. 246-251). In *Proceedings of the twenty-first annual conference of the cognitive science society*. Mahwah, NJ: Erlbaum.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kieras, D. E. & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12, 391-438.

Lazarus, R., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer Verlag.

Lebiere, C. (2001). A theory-based model of cognitive workload and its applications. In *Proceedings of the 2001 Interservice/Industry Training, Simulation and Education Conference* (I/ITSEC). Arlington, VA: NDIA

Lebiere, C., Anderson, J.R., & Bothell, D. (2001). Multi-tasking and cognitive workload in an ACT-R model of simplified air traffic control task. In *Proceedings of the 10th Conference on Computer-Generated Forces and Behavior Representation* (pp.91-98), May 15-21. Norfolk, VA.

Lebiere, C., Biefield, E., Archer, R., Archer, S., Allender, L., *et al*. (2002). IMPRINT/ACT-R: Integration of a task network modeling architecture with a cognitive architecture and its application to human error modeling.

Matthews, G. & Campbell, S.E. (1998). Task induced stress and individual differences in coping. In *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society* (pp.821-825). Santa Monica, CA: Human Factors and Ergonomics Society.

Matthews, G. (2001). Levels of transaction: A cognitive science framework for operator stress. In P.A. Hancock, & P.A. Desmond (Eds.), *Stress, workload, and fatigue.* Mahwah, NJ: Erlbaum.

O'Brien, L.H., Simon, R., & Swaminathan, H. (1992). Development of the personnel-based system evaluation aid (PER-SEVAL) performance shaping functions. *ARI Research Note 92-50*, United States Army Research Institute for the Behavioral and Social Sciences.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics,* vol.SMC-13, 257-266.

Sanders, A.F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, **53**, 61-97.

Smith, C.A.& Lazarus, R.S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion*, **7**, 233-269.

Ritter, F.E., Avraamides, M.N., Councill, I., Quigley, K.S., Klein, L.C. *et al.* (2002). Modeling the effects of two behavior moderators in ACT-R. Paper presented at the *Workshop on ACT-R Models of Human-System Interaction*, January 10th. Mesa, AZ.

Wickens C.D. (1980). The structure of attentional rersources. In R. Nickerson (Ed.). *Attention and performance VIII* (pp.239-257). Hillsdale, NJ: Erlbaum.

Wickens, C.D. (1996). Designing for stress. In J.E. Driskell & E. Salas (Eds.), *Stress and human performance* (pp.223-256). Mahwah, NJ: Lawrence Erlbaum.

Performance, Workload, and Stress with Temporal and Spatial Task Demands

J.M. Ross, J.L. Szalma, J. Thropp, & P.A. Hancock

Abstract

Increased understanding of the mechanisms by which stress impacts performance is essential to the design and operation of complex information systems. This study represents the initial test of the hypothesis by Hancock, Szalma, & Weaver (2002) that the attentional narrowing observed under stressful conditions results from spatial and temporal perception drawing on common resource capacities. Although the results were mixed regarding this hypothesis, a novel finding that noise induces leniency in responding was observed. This study reveals that the impact of noise stress on performance may depend on the characteristics of the task to be performed.

Introduction

The distortion of spatial perception under stress was conceived by Easterbrook (1959) as resulting from restriction in the range of cue utilization resulting from emotional arousal. This formulation was aimed at explaining the 'inverted-U' hypothesis regarding arousal (including stress) and performance. While the notion of unitary arousal has slowly dissolved, evidence for the narrowing of attention under stress has been experimentally validated (e.g. see Bacon, 1974; Cornsweet, 1969; Hancock & Dirkin, 1983; Dirkin & Hancock, 1984). Recently, Hancock and Weaver (2002) have argued that the narrowing in perception of time results from the same mechanism as narrowing of spatial perception, and that this narrowing is therefore a general effect of stress on perception of space-time. It was further argued by Hancock, Szalma, and Weaver (2002) that hypothetically this results from perception (and distortions of those perceptions) of spatial and temporal information drawing on common resource capacities. These narrowing effects should be exacerbated in the presence of an external stressor such as noise. In addition to performance effects, perceived workload and perceived stress level should increase when a task requires processing of spatial and temporal information. It has been well established that experimental manipulations that impact performance in target detection tasks also influence perceived workload and self-reports of stress in those tasks (e.g., Warm, Dember, & Hancock, 1996; Szalma, et al., in press). Performance decrements are often associated with increases in workload (although see Hancock & Parasuraman, 2001; Hancock, 1996), and contexts in which performance declines occur are associated with increases in self-reports of stress (e.g. Matthews et al., 2002; Szalma et al., in press). This experiment represents part of an on-going program to test hypotheses regarding the relation between changes in the perception of spatial and temporal processing under conditions of high stress and workload. Specifically, a target detection task was employed in which the relative strength of the spatial and temporal components was varied, such that these components were processed either separately or in combination. If spatial and temporal processing draw on similar resource capacities, performance decrements should be observed when the two components are combined as compared to tasks in which only one dimension is imposed.

Methods

Participants

Twenty-two females and twenty-two males, ranging in age from 18-32 (M = 21.3) at the University of Central Florida volunteered to participate.

Task

Three tasks were employed in the present study. The order of task performance was counterbalanced. In each task participants were required to discriminate a 2mm diameter O from a D and a "backwards-D," which were presented on a background mask of 4 mm diameter white circles on a black background (See Figure 1). The tasks differed according to spatial and temporal properties of target detection. The spatial-only task required observers to detect an O of brighter illumination that could appear in any one of nine locations on a screen (see Figure 1). Thus the spatial-only task presented observers with spatial uncertainty regarding target location. The 'temporal-only' required a temporal discrimination instead of an intensity discrimination, and observers were required to report when they observed an O of shorter duration (300 msecs). Non-targets were presentations of an O of longer duration (500 msecs), or a D or backwards D of either duration. In this condition however, there was no spatial uncertainty; stimuli were always presented in the center of the screen. In the third task ('combined') the temporal discrimination was combined with spatial uncertainty.

The event rate was 24 events per minute. In the conditions with spatial uncertainty the participant responded by pressing a key on the numerical keypad, from 1-9, corresponding to the quadrant on the screen in which the target appeared. In the temporal-only task, participants responded by pressing the five key on the numerical keypad. Correct responses occurring within 2.5 seconds of the onset of a signal were reported as Correct Detection; all other responses were recorded as False Alarms.

For the noise group an 85 dB intermittent white noise was presented via headphones during the three seven minute blocks. The intermittent white noise was composed of ten 4 second, eight 6 second, five 8 second, and four 10 second durations. The total duration of white noise per task was 2.8 minutes. Participants were randomly assigned to either the noise or quiet condition. Each participant experienced each of the three tasks, and was assigned at random to one of the six possible task orders.

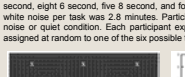


Figure 1. Perceptual sensitivity as a function of time for each task.

Questionnaires

Perceived workload was assessed using the NASA Task Load Index (TLX; Hart & Staveland, 1988), which provides an overall workload rating and ratings reflecting the relative contributions of six sources of workload: Mental Demand, Temporal Demand, Physical Demand, Performance, Effort, and Frustration. Self-reports of stress were measured using the Dundee Stress Questionnaire (DSSQ; Matthews et al, 2002). The DSSQ is composed of eleven scales reflecting the observer's cognitive and emotional states. These include Tense Arousal, Hedonic Tone, and Confidence and Control, reflecting the individual's degree of distress; Energetic Arousal, Concentration, and two kinds of motivation (Intrinsic and Success) which indicate the individual's degree of task-engagement; and Self-Esteem, Self-Focused Attention, and two forms of cognitive interference (Task-Irrelevant and Task-Related) which reflect the individual's degree of worry.

Procedure

Following the instructions for the first task participants completed the pre-DSSQ, after which they began the first task. Upon completion of each task participants were administered the post-DSSQ and the NASA-TLX, which were administered in a randomly assigned counterbalanced order. This order was repeated after the second and third tasks. Upon completion participants were debriefed and thanked for their participation.

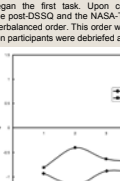


Figure 3. Response bias as a function of time on task for each noise condition.



Figure 4. Overall Workload as a Function of Task Type and Noise Condition.

Results

Performance. Perceptual sensitivity increased over the first three minutes and declined over the last three minutes of performance in each task (See Figure 2). In addition, sensitivity in the temporal-only task was higher than the sensitivity scores in the other two tasks. These impressions were confirmed by an ANOVA, which revealed a significant effect for task, $F(2,68)=6.23$, $p<.01$, and for time, $F(4,181)=2.41$, $p=.05$. All other effects failed to reach significance ($p>.05$ in each case). Post-hoc comparisons revealed that observers were significantly more sensitive in the temporal-only task than in the spatial-only task and the combined task, but that their scores in the latter two tasks did not significantly differ from one another.

An ANOVA of the (c) scores revealed that observers in the noise condition were significantly more lenient in responding than their cohorts in the quiet condition (see Figure 3), $F(1,42)=4.98$, $p=.03$. All other effects failed to reach significance ($p>.05$ in each case).

Perceived Workload. Preliminary analysis revealed that scores on the NASA-TLX did not significantly vary as a function of the order of task performance. An ANOVA of the Global workload ratings revealed a significant main effect for task, $F(2, 82) = 3.53$, $p < .05$, and a marginally significant effect for noise, $F(1, 42) = 3.57$, $p = .07$. Post-hoc comparisons indicated that the temporal task induced significantly less global workload than the other two tasks, which did not significantly differ from one another. In addition to global workload, weighted ratings on the subscales of the TLX were obtained. An ANOVA revealed a significant effect for noise, $F(1, 42) = 4.30$, $p = .04$, with the noise condition inducing significantly higher workload than the quiet condition. Significant effects were also observed for scales, $F(3, 124) = 9.72$, $p = .00$, and a task by scales interaction, $F(6.242) = 2.45$, $p = .03$. Separate task by noise ANOVAS indicated significant effects for the Weighted Temporal Demand subscale, $F(2, 83) = 5.90$, $p < .01$, and post-hoc comparisons indicated that the combined task induced more temporal demand than either of the other two tasks, which did not significantly differ from one another (See Figure 4). A significant effect for task was also observed in the Weighted Performance subscale, $F(2, 78) = 3.61$, $p < .05$, and post-hoc tests revealed that the combined task induced significantly less Performance Workload than the spatial-only task. A marginal effect for noise was observed in the Weighted Frustration subscale, $F(1, 42) = 3.91$, $p = .06$, with noise having higher frustration levels than quiet conditions.

Stress. Participant's responses on the pre- and post-DSSQ were used to calculate scores on eleven scales using the formula (post-score – pre-score/SD, where the standard deviation use was obtained from a large normative group obtained from Matthews et al. (1999), except the two motivation scales, for which normative values are not available. For these scales, sample standard deviations were used (Matthews, Campbell, & Falconer, 2001). Separate 2 (Noise) by 3 (Task) ANOVAs were computed for each scale. Note that preliminary tests revealed that the order of task performance did not exert a significant influence on DSSQ change scores. Significant task effects were observed for Energetic Arousal, $F(2, 75) = 5.82$, $p = .006$, and post-hoc tests indicated that the temporal-only task induced a significantly larger drop in Energetic Arousal relative to pre-task state as compared to the spatial only and combined tasks, which did not significantly differ from one another (See Figure 6). Tense Arousal had a

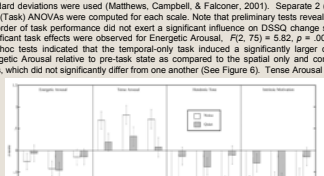


Figure 6. DSSQ Scale Scores as a Function of Task Type and Noise Condition. Note. Error bars are Standard Error.

significant main effect for noise, $F(1, 42) = 3.98$, $p = .05$. Observers in the noise condition reported significantly higher Tense Arousal than observers in the quiet condition. A significant noise effect was also observed for the Hedonic Tone scale, $F(1, 42) = 6.84$, $p = .01$. Participants in the noise condition showed a greater decline in Hedonic Tone than observers in the quiet condition. Further, a significant main effect for task was observed, $F(2, 83) = 3.07$, $p = .05$. Post-hoc tests revealed that hedonic tone was significantly lower after a temporal task than in the combined task, but that change scores after the spatial-only task did not differ significantly from those associated with the other two tasks. For Intrinsic Motivation there was a marginally significant main effect for task, $F(2, 72) = 2.96$, $p = .07$, with temporal-only task tending to show greater declines in Intrinsic Motivation than those in the combined task .

Discussion

Due to potential ceiling effects in the detection rate itself (average d' scores were above 5 for all three tasks), the performance data did not provide unequivocal information either for or against the hypothesis of Hancock et al. (2002) regarding the covariation of temporal and spatial distortion. These effects extended to perceived workload and two scales of the DSSQ (Energetic Arousal and Intrinsic Motivation). It may be that differences in performance would be observed with more difficult discriminations, and efforts are currently underway to investigate this possibility.

However, two subscales of the TLX and three subscales of the DSSQ indicated that task type had effects on subjective ratings that differed from those described above. Thus, results with the Weighted Temporal Demand scale showed that while spatial uncertainty (spatial-only task) and temporal discrimination without uncertainty (temporal-only task) induced similar Weighted Temporal Demand, combining these features (combined task) increased Weighted Temporal Demand. In contrast, participants in the combined condition rated their performance as better in comparison to the spatial-only task, indicating a facilitative effect of adding a temporal discrimination to a task with spatial uncertainty. The latter effect accords with the main effect for task observed in the Hedonic Tone scale of the DSSQ, in which observers reported lower Hedonic Tone after the temporal-only task than after the combined task. Contrary to expectations, combining the temporal and spatial elements of a task reduced the effects of task stress and workload along these dimensions.

In this study noise did not impact sensitivity, but it did decrease response bias, indicating more lenient responding in the presence of noise. A marginal noise by time interaction suggested a rise in conservatism in the quiet condition and no change over time in the noise condition. These results are consistent with the general view that stress lowers response criteria (Welford, 1973), but not with the literature on the impact of noise, which indicates that the criterion should increase (e.g. see Broadbent, 1971) or not significantly change (e.g. Milosevic, 1983). The consensus, based upon the work of Broadbent (1971; see also Poulton, 1977), is that in long duration tasks continuous noise induces a rise in conservatism, particularly for responses of intermediate confidence, and that in shorter duration tasks noise may serve to enhance performance. However, as noted in this study no performance enhancement (in terms of d') was observed, and response bias decreased. Note that in a short duration vigil using similar stimuli to those employed in this study, (Helton, Warm, Matthews, Corcoran, & Dember, 2002) continuous jet engine noise improved detections without changing the false alarm rate. A crucial difference between the study of Helton and his colleagues and the current study was the format of noise. In their study, noise was continuous presentation of jet engine noise, while in the current study intermittent noise of variable duration and random presentation times was used. Thus, participants experienced a greater degree of uncertainty regarding when and for how long noise would be present. It may be that it is uncertainty as a source of stress that induces leniency in responding.

The impact of noise on observers' state was more pronounced in workload and stress, with participants experiencing noise reporting increases in Weighted Frustration and a marginal increase in Global Workload. Similarly, participants in the noise condition reported more Tense Arousal and lower Hedonic Tone, and in the case of conditions with no spatial uncertainty, less Intrinsic Motivation. These effects differ from those of Helton and his colleagues, who observed that noise decreased the symptoms of stress. In their study, noise decreased the anger/frustration experienced by observes after a difficult vigilance task, and increased energetic arousal, intrinsic motivation and confidence and control relative to a quiet condition. The findings in regard to the latter three scales indicate an increase in task engagement in the presence of noise. These results represent trends opposite to those observed in this study, despite the fact that the task demands placed upon operators by Helton et al. were substantially greater than those in this study. It may be that whether noise exerts a negative effect on stress and performance depends on how the difficulty level of the task interacts with the noise presented to operators. Future research will investigate this possibility.

References

Bacon, S.J., (1974). Arousal and the range of cue utilization. Journal of Experimental Psychology, 102(1), 81-87.

Broadbent, D.E. (1971). Decision and stress. London: Academic Press.

Dirkin, G.R., & Hancock, P.A. (1984). Attentional narrowing to the visual periphery under hypoxic and anoxic stress. Aviation, Space and Environmental Medicine, 55, 457 (abstract only).

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 66, 183-201.

Hancock, P.A. (1996). Effect of control order, augmented feedback, input device and practice on tracking performance and perceived workload. Ergonomics, 39, 1146-1162.

Hancock, P.A., & Dirkin, G.R. (1983). Stressor induced attentional narrowing: Implications for design and operation of person-machine systems. Proceedings of the Human Factors Association of Canada, 15, 63-67.

Hancock, P.A., Szalma, J.L., & Weaver, J.L. (2002). Distortion of perceptual space-time under stress. Paper presented at the 23rd Army Science Conference, Orlando, FL, December, 2002.

Hancock, P.A., & Weaver, J.L. (in press). On time distortion under stress. Theoretical Issues in Ergonomic Science.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp139-183). Amsterdam: North-Holland.

Helton, W.S., Warm, J.S., Matthews, G., Corcoran, K.J., & Dember, W.N. (2002). Further tests of an abbreviated vigilance task: Effects of signal salience and jet aircraft noise on performance and stress. Proceedings of the Human Factors and Ergonomics Society, 46, 1949-1953.

Matthews, G., Campbell, S.E., & Falconer, S., (2001). Assessment of motivational states in performance environments. Proceedings of the Human Factors and Ergonomics Society, 45, 906-910.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J., Gilliland, K. et al. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. Emotion, 2(4), 315-340.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards the state "big three"? In I. Mervielde, I.J. Deary, F. De Fruyt, & F. Ostendorf (Eds), Personality psychology in Europe (vol 7) (pp. 335-350).

Milosevic, S., (1982). Effects of noise on signal detection. Ergonomics, 26(12), 929-936.

Poulton, E.C., (1977). Continuous intense noise masks auditory feedback and inner speech. Psychological Bulletin, 84(5), 977-1001.

Poulton, E.C., (1977). Continuous noise and shifts of criterion. Psychological Bulletin, 84(5), 901-907.

Szalma, J.L. (in press). Workload and stress of configural displays in vigilance tasks. Proceedings of the Human Factors and Ergonomics Society, 48.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In R. Parasuraman and M. Mouloua. (Eds). Automation and human performance: Theory and applications. (pp.183-200). Erlbaum, Hillsdale, NJ.

Welford, A.T., (1973). Stress and performance. Ergonomics, 16, 567-580.

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office. P.A. Hancock, principal investigator. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Dillingas for providing administrative and technical direction for the Grant.

PERFORMANCE, WORKLOAD, AND STRESS WITH TEMPORAL AND SPATIAL TASK DEMANDS

Ross, J.M, Szalma, J.L., Thropp, J., & Hancock, P.A

University of Central Florida

ABSTRACT

Increased understanding of the mechanisms by which stress impacts performance is essential to the design and operation of complex information systems. This study represents the initial test of the hypothesis by Hancock, Szalma, & Weaver (2002) that the attentional narrowing observed under stressful conditions results from spatial and temporal perception drawing on common resource capacities. Although the results were mixed regarding this hypothesis, a novel finding that noise increases leniency in responding observed. This study reveals that the impact of noise stress on performance may depend on the characteristics of the task to be performed, with spatial uncertainty exerting an influence on stress and perceived workload.

SUMMARY

The distortion of spatial perception under stress was conceived by Easterbrook (1959) as resulting from restriction in the range of cue utilization resulting from emotional arousal. This formulation was aimed at explaining the 'inverted-U' hypothesis regarding arousal (including stress) and performance. While the notion of unitary arousal has slowly dissolved, evidence for the narrowing of attention under stress has been experimentally validated (e.g. for a review see Hancock, Szalma, & Weaver, 2002). Recently, Hancock and Weaver (2002) have argued that the narrowing in perception of time results from the same mechanism as narrowing of spatial perception, and that this narrowing is therefore a general effect of stress on perception of space-time. It was further argued by Hancock,

Poster presented at the Mid-Year Symposium of Division 21 and Division 19 of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society: Contemporary and Emerging Issues in Human Factors, Engineering and Military Psychology, Ft. Belvoir, VA, March, 2003

Szalma, and Weaver (2002) that hypothetically this results from perception (and distortions of those perceptions) of spatial and temporal information drawing on common resource capacities. These narrowing effects should be exacerbated in the presence of an external stressor such as noise. In addition to performance effects, perceived workload and perceived stress level should increase when a task requires simultaneous processing of spatial and temporal information. It has been well established that experimental manipulations that impact performance in target detection tasks also influence perceived workload and self-reports of stress in those tasks (e.g., Warm, Dember, & Hancock, 1996; Szalma, et al., in press). This experiment represents part of an on-going program to test hypotheses regarding the relation between changes in the perception of spatial and temporal processing under conditions of high stress and workload. Specifically, a target detection task was employed in which the relative strength of the spatial and temporal components was varied, such that these components were processed either separately or in combination. If spatial and temporal processing draw on similar resource capacities, performance decrements should be observed when the two components are combined as compared to tasks in which only one dimension is imposed.

Twenty-two females and twenty-two males, at the University of Central Florida volunteered to participate. Participants ranged in age from 18 to 32, with a mean age of 21.27 years (SD = 3.28).

Three tasks were employed in the present study, with the order of task performance counterbalanced across participants. Each task required discrimination of a 2mm diameter O from a D and a "backwards-D," which were presented on a background mask of .4 mm diameter white circles on a black background. The tasks differed

according to spatial and temporal properties of target detection. The spatial-only task required observers to detect an O of brighter illumination that could appear in any one of nine locations on a screen. Thus the spatial-only task presented observers with spatial uncertainty regarding target location. The 'temporal-only' required a temporal discrimination instead of an intensity discrimination, and observers were required to report when they observed an O of shorter duration (300 msecs). Non-targets were presentations of an O of longer duration (500 msecs), or a D or backwards D of either duration. In this condition however, there was no spatial uncertainty; stimuli were always presented in the center of the screen. In the third task ('combined') the temporal discrimination was combined with spatial uncertainty. Thus, observers were required to detect an O of shorter duration that could appear in any one of nine locations on the screen. Note that the number of conjunction searches was constant across the three tasks.

The event rate was 24 events per minute. In the conditions with spatial uncertainty the participant responded by pressing a key on the numerical keypad, from 1-9, corresponding to the quadrant on the screen in which the target appeared. In the temporal-only task, participants responded by pressing the five key on the numerical keypad. Correct responses occurring within 2.5 seconds of the onset of a signal were reported as Correct Detection; all other responses were recorded as False Alarms.

For the noise group 85 dB intermittent white noise was presented via headphones during the three seven minute blocks. The intermittent white noise was composed of ten 4 second, eight 6 second, five 8 second, and four 10 second durations. The total duration of white noise per task was 2.8 minutes. Observers in the quiet condition also wore headphones to control for the effects of additional stimulation. The ambient sound level

in the quiet condition was 64.5 dB. Participants were randomly assigned to either the noise or quiet condition.

Perceived workload was assessed using the NASA Task Load Index (TLX; Hart & Staveland, 1988), which provides an overall workload rating and ratings reflecting the relative contributions of six sources of workload: Mental Demand, Temporal Demand, Physical Demand, Performance, Effort, and Frustration. Self-reports of stress were measured using the Dundee Stress Questionnaire (DSSQ; Matthews et al., 2002). The DSSQ is composed of eleven scales reflecting the observer's cognitive and emotional states. These include Tense Arousal, Hedonic Tone, and Confidence and Control, reflecting the individual's degree of Distress; Energetic Arousal, Concentration, and two kinds of motivation (Intrinsic and Success) which indicate the individual's degree of Task-Engagement; and Self-Esteem, Self-Focused Attention, and two forms of cognitive interference (Task-Irrelevant and Task-Related) which reflect the individual's degree of Worry.

Participants completed the pre-DSSQ prior to engaging in the first task. After each task participants were administered the post-DSSQ and the NASA-TLX, which were administered in a randomly assigned counterbalanced order. This order was repeated after the second and third tasks.

The correct detection and false alarm rates for each participant were used to compute signal detection theory measures of sensitivity (d') and response bias (c) in each minute of task performance. An ANOVA revealed a significant effect for task, $F(2,68)=6.23$, $p<.01$, was observed, and a significant effect for time on task was obtained, $F(4,181)=2.41$, $p=.05$. All other effects failed to reach significance ($p>.05$ in

each case). Note that in all test involving repeated measures factors Box's epsilon was used to adjust the degrees of freedom for violations of the sphericity assumption (Myers & Well, 1995). Post-hoc comparisons revealed that observers were significantly more sensitive in the temporal-only task than in the spatial-only task and the combined task, but that their scores in the latter two tasks did not significantly differ from one another.

Mean response bias scores for each noise condition are plotted as a function of time on task in Figure 2. An ANOVA of the (c) scores revealed that observers in the noise condition (M=-.93) were significantly more lenient in responding than their cohorts in the quiet condition (M=-.43), $F(1,42)=4.98$, $p=.03$. A marginal effect for time was also observed, $F(4,173)=2.13$, $p=.08$, with a trend for conservatism to rise over time. All other effects failed to reach significance ($p>.05$ in each case).

Preliminary analysis revealed that scores on the NASA-TLX did not significantly vary as a function of the order of task performance. Global workload ratings for the NASA-TLX were calculated and revealed that as a main effect, the temporal-only task induced the least global workload (See Figure 3). Further, noise induced higher workload in each task. An ANOVA confirmed these impressions, with a significant main effect for task, $F(2, 82) = 3.53$, $p < .05$, and a marginally significant effect for noise, $F(1, 42) = 3.57$, $p = .07$. Post-hoc comparisons indicated that the temporal task induced significantly less global workload than the other two tasks, which did not significantly differ from one another. In addition to global workload, weighted ratings on the subscales of the TLX were obtained. An ANOVA revealed a significant effect for noise, $F(1, 42) = 4.30$, $p = .04$, with the noise condition inducing significantly higher workload (M = 167.71) than the quiet condition (M = 140.97). Significant effects were also observed for

scales, $F(3, 124) = 9.72$, $p = .0001$, and a task by scales interaction, $F(6,242) = 2.45$, $p = .03$. Separate task by noise ANOVAS for each scale were computed. Weighted ratings for three of the subscales are shown as a function of task-type and noise condition in Figure 6. Significant effects for task were observed for the Weighted Temporal Demand subscale, $F(2, 83) = 5.90$, $p < .01$, and post-hoc comparisons indicated that the combined task induced more temporal demand than either of the other two tasks, which did not significantly differ from one another. In addition, a marginal noise by task interaction was obtained, $F(2, 83) = 2.67$, $p = .08$, with a trend for noise to induce more temporal demand in the spatial-only task (see Figure 4). A significant effect for task was also observed in the Weighted Performance subscale, $F(2, 78) = 3.61$, $p < .05$, and post-hoc tests revealed that the combined task induced significantly less Performance Workload than the spatial-only task. A marginal effect for noise was observed in the Weighted Frustration subscale, $F(1, 42) = 3.91$, $p = .06$. with noise having higher frustration levels (M = 166.82) than quiet conditions (M = 91.30).

Participant's responses on the pre- and post-DSSQ were used to calculate scores on eleven scales using the formula (post-score – pre-score/SD, where the standard deviation use was obtained from a large normative group obtained from Matthews et al. (1999). Separate 2 (Noise) by 3 (Task) ANOVAs were computed for each scale. Note that preliminary tests revealed that the order of task performance did not exert a significant influence on DSSQ change scores. Significant task effects were observed for Energetic Arousal, $F(2, 75) = 5.82$, $p = .006$, and post-hoc tests indicated that the temporal-only task (M=-.44) induced a significantly larger drop in Energetic Arousal relative to pre-task state as compared to the spatial only (M=-.16) and combined (M=-

.15) tasks, which did not significantly differ from one another. Tense Arousal had a significant main effect for noise, $F(1, 42) = 3.98$, $p = .05$. Observers in the noise condition reported significantly higher Tense Arousal (M = 0.74) than observers in the quiet condition (M = 0.20). A significant noise effect was also observed for the Hedonic Tone scale, $F(1, 42) = 6.84$, $p = .01$. Participants in the noise condition showed a decline in Hedonic Tone (M = -0.90) than observers in the quiet condition (M = -0.20). Further, a significant main effect for task was observed, $F(2, 83) = 3.07$, $p = .05$. Post-hoc tests revealed that hedonic tone was significantly lower after a temporal task (M = -0.67) than in the combined task (M = -0.46), but that change scores after the spatial-only task (M=-.52) did not differ significantly from those associated with the other two tasks. For Intrinsic Motivation there was a marginally significant main effect for task, $F(2, 72) = 2.96$, $p = .07$, with temporal-only task (M = -0.67) tending to show greater declines in Intrinsic Motivation than those in the combined task (M = -0.38).

Due to potential ceiling effects in the detection rate itself (d' scores above 5 for all three tasks), the performance data did not provide unequivocal information either for or against the hypothesis of Hancock et al. (2002) regarding the covariation of temporal and spatial distortion. Thus the only difference in sensitivity was observed in the expected superiority of the task without spatial uncertainty. These effects extended to perceived workload, which was higher in tasks with spatial uncertainty, regardless of whether one of the discriminations was temporal or non-temporal. Similarly, both tasks with spatial uncertainty reduced Energetic Arousal (with a marginally significant trend for Intrinsic Motivation) relative to the temporal task without such uncertainty. It may be that

differences in performance would be obtained with more difficult discriminations, and efforts are currently underway to investigate this possibility

However, two subscales of the TLX and three subscales of the DSSQ indicated that task type had effects that differed from those described above. Thus, results with the Weighted Temporal Demand scale showed that while spatial uncertainty (spatial-only task) and temporal discrimination without uncertainty (temporal-only task) induced similar Weighted Temporal Demand, combining these features (combined task) increased Weighted Temporal Demand, although a marginal interaction effect between task and noise suggested that this effect was stronger in the quiet condition than in the noise condition. In contrast, participants in the combined condition rated their performance as better in comparison to the spatial-only task, indicating a facilitative effect of adding a temporal discrimination to a task with spatial uncertainty. The latter effect accords with the main effect for task observed in the Hedonic Tone scale of the DSSQ, in which observers reported lower hedonic tone after the temporal-only task than after the combined task, regardless of the order in which the tasks were performed. Similarly, a marginal effect for intrinsic motivation indicated that the temporal-only task induced greater reductions in intrinsic motivation relative to the combined task. Contrary to expectations, combining the temporal and spatial elements of a task reduced the effects of task stress along these dimensions. Current experimental work will investigate whether these effects extend to a task in which spatial and temporal discriminations are combined rather than temporal discrimination and spatial uncertainty.

In this study noise did not impact sensitivity, but it did decrease response bias, indicating more lenient responding in the presence of noise. These results are consistent with the

general view that stress lowers reponse criteria (Welford, 1973), but not with the literatiure on the impact of noise, which indicates that the criterion should increase (e.g., see Broadbent, 1971) or not significantly change (e.g., Harper, 1979; Milosevic, 1983). The consensus, based upon the work of Broadbent (1971; 1978; see also Poulton, 1977; 1978), is that in long duration tasks continuous noise induces a rise in conservatism, particularly for responses of intermediate confidence, and that in shorter duration tasks noise may serve to enhance performance. However, in this study, which employed a short duration task, no performance enhancement (in terms of d') was observed, and response bias decreased. Note that in a short duration vigil using similar stimuli to those employed in this study (Helton, Warm, Matthews, Corcoran, & Dember, 2002) continuous jet engine noise improved detections without modifying the false alarm rate. A crucial difference between the study of Helton et al. and the current study was the format of noise. They employed continuous of jet engine noise, while in the current study intermittent noise of variable duration and random presentation times was used. Thus, participants experienced a greater degree of uncertainty regarding when and for how long noise would be present. It may be that it is uncertainty as a source of stress that induces leniency in responding.

The impact of noise on observers' state was revealed in self-reports of workload and stress, with participants experiencing noise reporting increases in Weighted Frustration and a marginal increase in Global Workload. Similarly, participants in the noise condition reported more Tense Arousal and lower Hedonic Tone, and in the case of conditions with no spatial uncertainty, less Intrinsic Motivation. These effects differ from those of Helton et al., who observed that noise decreased the symptoms of stress by

reducing Frustration and increasing Task Engagement. These results represent trends opposite to those observed in this study, despite the fact that the task demands placed upon operators by Helton et al. were substantially greater than those in this experiment. For instance, the event rate in their study was 57.5 events per min, while an event rate of 24 events per minute was employed herein. It may be that whether noise exerts a negative effect on stress and performance depends on how the difficulty level of the task interacts with the noise presented to operators. Future efforts will investigate this possibility.

References

Bacon, S.J., (1974). Arousal and the range of cue utilization. Journal of Experimental Psychology, 102(1), 81-87.

Broadbent, D.E., (1971). Decisions and Stress. England: Medical Research Council.

Broadbent, D.E., (1978). The current state of noise research: Reply to Poulton. Psychological Bulletin, 85(5), 1052-1067.

Dirkin, G.R., & Hancock, P.A. (1984). Attentional narrowing to the visual periphery under temporal and acoustic stress. Aviation, Space and Environmental Medicine, 55, 457 (abstract only).

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 66, 183-201.

Hancock, P.A., & Dirkin, G.R. (1983). Stressor induced attentional narrowing: Implications for design and operation of person-machine systems. Proceedings of the Human Factors Association of Canada, 16, 19-21.

Hancock, P.A., Szalma, J.L, & Weaver, J.L. (2002). Distortion of perceptual space-time under stress. Paper presented at the 23rd Army Sciences Conference, Orlando, FL, December, 2002.

Hancock, P.A., & Weaver, J.L. (in press). On time distortion under stress. Theoretical Issues in Ergonomic Science.

Harper, D.W., (1979). Signal detection analysis of effect of white noise intensity on sensitivity to visual flicker. Perceptual & Motor Skills, 48(3), 791-798.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.). Human mental workload (pp. 139-183). Amsterdam: North-Holland.

Helton, W.S., Warm, J.S., Matthews, G., Corcoran, K.J., & Dember, W.N. (2002). Further tests of an abbreviated vigilance task: Effects of signal salience and jet aircraft noise on performance and stress. Proceedings of the Human Factors and Ergonomics Society, 46, 1546-1550.

Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., et al. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. Emotion, 2(4), 315-340.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards the state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds), Personality psychology in Europe (vol 7), (pp. 335-350).

Milosevic, S., (1983). Effects of noise on signal detection. Ergonomics, 26(10), 939-946.

Myers, J.L., & Well, A.D. (1995). Research Design and Statistical Analysis. (2nd ed.). New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Poulton, E.C., (1977). Continuous intense noise masks auditory feedback and inner speech. Psychological Bulletin, 84(7), 977-1001.

Poulton, E.C., (1978). A new look at the effects of noise: A rejoinder. Psychological Bulletin, 85(5), 1068-1079.

Temple, J.G., Warm, J.S., Dember, W.N., Jones, K.S., LaGrange, C.M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. Human Factors, 42, 183-194.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In: R. Parasuraman and M. Mouloua. (Eds). Automation and human performance: Theory and applications. (pp 183-200), Erlbaum, Hillsdale, N.J.

Welford, A.T., (1973). Stress and performance. Ergonomics, 16, 567-580.

Comparison of Methods of Analyses from Fuzzy & Traditional Signal Detection Theory

L.L. Murphy, J.L. Szalma, P.A. Hancock, & S.C. Stafford

Abstract

A recent advance in Signal Detection Theory (SDT) promises to enhance measurement of performance in complex real world domains where stimuli do not fall into discrete, mutually exclusive categories.. This development, Fuzzy Signal Detection Theory (FSDT; Parasuraman, Masalonis, & Hancock, 2000), combines traditional SDT with Fuzzy Set Theory to extend signal detection analysis beyond the traditional crisp, categorical decision making model. FSDT allows for events to simultaneously be in more than one state category (e.g. signal and non-signal), so that the stimulus and response dimensions can be continuous rather than categorical. This study compared the differences in methods of analyses from FSDT and traditional SDT using the same data set.

Introduction

For almost fifty years Signal Detection Theory (SDT) has provided an extremely useful analytical tool for evaluating human and machine performance in both simple and complex systems. As such, it has served as one of the most useful measurement tools and influential theories in psychology. However, even outstanding theories have limitations, and SDT is no exception. In the SDT model, the state of the world is forced into crisp mutually exclusive categories (i.e., signal versus non-signal; old versus new; etc.), which may not be accurate representations of the true states of the world. To capture the uncertainty inherent in real-world stimuli, the representation of the stimulus dimension must be continuous rather than categorical. Recently, Parasuraman, Masalonis, and Hancock (2000; see also Hancock, Masalonis, & Parasuraman, 2000) have captured this uncertainty by combining elements of SDT with those of Fuzzy Set Theory, in which category membership is not considered mutually exclusive and stimuli can therefore be simultaneously assigned to more than one category. Thus, in FSDT a given stimulus, or more formally, a stimulus selected at random from a distribution of stimuli, may be categorized as both a correct detection and a false alarm depending on the degree to which the stimulus represents a critical event. For instance, a convenient range for a stimulus dimension is one in which the strength of the stimulus varies from 0 (a definite non-signal) and 1 (definitely a signal), with a signal value of .5 representing maximal uncertainty in the stimulus itself. That is, a stimulus with a signal value of .5 has properties of both a non-signal and a signal. Implicit in this model is the assumption that signal uncertainty exists not only within the observer (a major insight provided by traditional SDT) but in the stimulus dimension itself. This experiment represents an initial test of FSDT using a perceptual signal detection task. In this study fuzzy stimulus and response dimensions were established (seven categories for each), and difficulty (temporal) and bias (instructional set) were manipulated to investigate the degree to which the FSDT model is sensitive to these classical SDT manipulations. The FSDT analysis was compared to an analysis using traditional SDT, using methods outlined in MacMillan and Creelman (1991).

Method

Participants

Nine students from the University of Central Florida (3 men and 6 women, mean age = 21 years) volunteered to participate in this study. Volunteers were compensated by extra credit in their psychology course and were told that a $25 award will be given to the two best performers.

Stimuli

In each trial, the event rate of the stimuli was 21 event/min. Each of the 7 different stimuli was presented 18 times. The order in which the stimuli were presented was randomized. Note that participants were not informed of the number of stimulus categories or the frequency distribution of the stimuli.

Procedure

A computerized target detection task, that required color discrimination, was used to collect performance data (d' and C)

After instructions, participants completed a practice trial of the task. Both difficulty levels of the task (less difficult 300 ms stimulus duration) vs. more difficult (150 ms stimulus duration) were presented in the practice session, the order in which the participants received the difficulty levels was counterbalanced.

Participants engaged in six 6-min detection tasks consisting of two difficulty levels and three levels of instruction bias (lenient, conservative and neutral). The order in which the participants received the different treatments was counterbalanced using a Latin Square design.

MIT² INSTITUTE for SIMULATION & TRAINING

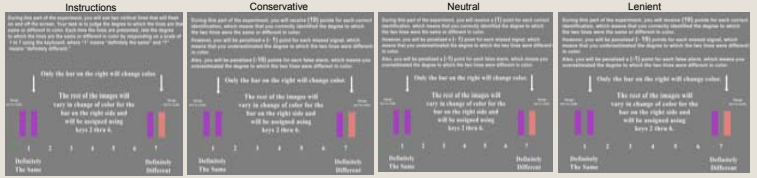
University of Central Florida

Multi-Disciplinary University Research Initiative

MURI-OPUS

Operator Performance Under Stress

www.mit.ucf.edu



Results

A 2 x 3 repeated measures analysis of variance was the statistical test used to analyze the data. The independent variables were difficulty level (less vs. more difficult) and bias (lenient, neutral, and conservative). Dependent variables were scores of d' (a measure of perceptual sensitivity) and C (a measure of bias) calculated using both Fuzzy Signal Detection and Traditional Signal Detection Theories of analyses from the same data set.

Figure 1 reveals that when d' was calculated using the Traditional Signal Detection method of analysis collapsed across stimuli, participants were significantly more sensitive in color discrimination for the less difficult stimuli compared to the more difficult stimuli, F (1, 12) = 9.65, p = .009.

Additionally, the data revealed that participants significantly differed in their d' scores across bias conditions, F (2, 19) = 10.16, p = .002. Figure 2 reveals that participants' sensitivity for color discrimination was greatest in the lenient condition, followed by the neutral condition. The conservative condition had the least amount of sensitivity.

In comparison, when d' was calculated using the Fuzzy Signal Detection method of analysis collapsed across stimuli, the data revealed that participants marginally significantly differed in their d' scores across bias conditions, F (2, 9) = 3.91, p = .06. The trend suggest that sensitivity is greater in the conservative condition (M = 0.99, SD = 0.35) compared to the lenient condition (M = 0.44, SD = 0.19).



Figure 1. Traditional SDT method of analysis collapsed across stimuli for d' and difficulty.

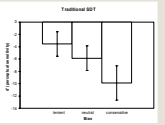


Figure 2. Traditional SDT method of analysis collapsed across stimuli for d' and bias.

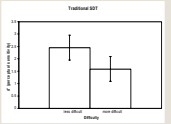


Figure 3. Traditional SDT method of analysis by stimulus for d' and difficulty.

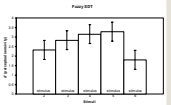


Figure 4. Fuzzy SDT method of analysis by stimulus for d' and stimulus.

Figure 3 reveals that when d' was calculated using the Traditional Signal Detection method of analysis by stimuli, participants were significantly more sensitive in color discrimination for the less difficult stimuli compared to the more difficult stimuli, F (1, 12) = 9.65, p = .009. In contrast, when d' was calculated using the Fuzzy Signal Detection method of analysis, the data revealed no significant effect for difficulty level.

When d' was calculated using the Fuzzy Signal Detection method of analysis by stimuli, the data revealed that participants color discrimination was significantly different across stimuli (2 through 6), F (3, 31) = 15.58, p < .0005.

Figure 4 reveals that participants' sensitivity for color discrimination was significantly greater for stimuli with maximal uncertainty (stimuli 3, 4, and 5) and lower for stimuli with minimal uncertainty (stimuli 2 and 6). This trend appears to be stronger for the less difficult stimuli compared to the more difficult stimuli. In contrast, when d' was calculated using the Traditional Signal Detection method of analysis by stimuli, the data revealed no significant effect for stimuli.

When C was calculated using the Fuzzy Signal Detection method of analysis by stimuli, the data revealed a significant effect for stimuli (2 through 6), F (2, 23) = 27.96, p < .0005. Figure 6 reveals that participants' bias (degree of conservativeness and leniency) in color discrimination was significantly different from stimuli with the least uncertainty (stimuli 2 and 6) compared to the stimuli with the most uncertainty (stimuli 3, 4, and 5). Data revealed that participants responded to stimulus 6 with the highest bias toward a lenient, followed by stimulus 5, stimulus 4, and stimulus 3. Participants responded to stimulus 2 with the highest bias toward a conservative response.

Similarly, when C was calculated using the Traditional Signal Detection method of analysis by stimuli, the data revealed a significant interaction for difficulty (less difficult vs. more difficult) by stimuli (stimuli 1 to 2, 2 to 3, 3 to 4, 4 to 5, 5 to 6, and 6 to 7), F (4, 38) = 5.23, p = .003. Figure 7 reveals that participants' bias (degree of conservativeness and leniency) in color discrimination was significantly different for stimuli presented in the less difficult condition compared to the more difficult condition. Data revealed that participants responded to stimuli 5 to 6 with the highest bias toward a lenient response, followed by stimuli 4 to 5, stimuli 3 to 4, stimuli 6 to 7, and stimuli 2 to 3. Participants responded to stimuli 1 to 2 with the highest bias toward a conservative response (M = -1.88, SD = 0.52).

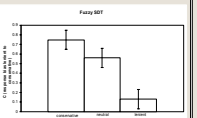


Figure 5. Fuzzy SDT method of analysis by stimulus for three levels of c.

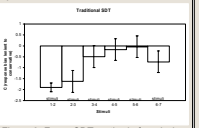


Figure 6. Fuzzy SDT method of analysis by stimulus for three levels of c.

When C was calculated using the Fuzzy Signal Detection method of analysis by stimuli, the data revealed a significant effect for bias (conservative, neutral, and lenient), F (2, 23) = 13.76, p < .0005. Figure 5 reveals that participants' bias (degree of conservativeness and leniency) in color discrimination was significantly more lenient for stimuli presented in the lenient condition compared to the participants' bias toward stimuli presented in the conservative and neutral conditions, which did not significantly differ from each other.

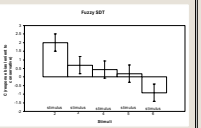
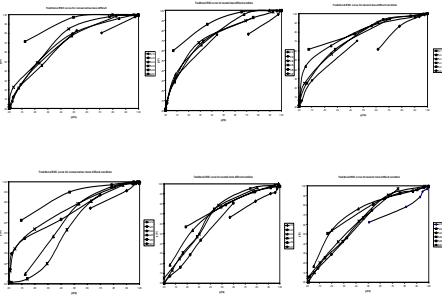
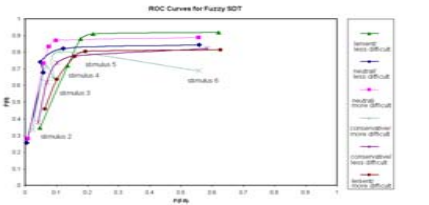


Figure 7. Traditional SDT method of analysis by stimulus for C and stimulus in the less difficult condition.

ROC Curves for Traditional SDT



ROC Curves for Fuzzy SDT



Conclusions

Fuzzy Signal Detection Theory analysis and Traditional Signal Detection Theory provide different vistas into signal detection performance. That FSDT provided a better description of the effects of stimulus uncertainty on observers' responses is not surprising, given that the FSDT model explicitly captures this uncertainty. However, the finding that the traditional model is more sensitive to the difficulty manipulation suggest that further work is necessary to investigate the relation between task difficulty and the FSDT model. These results do indicate, however, that FSDT can provide insight into system performance in domains in which stimulus categories varly along a continuum.

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, P.A. Hancock, principal investigator. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

University of Central Florida

Multi-Disciplinary University Research Initiative

MURI-OPUS

Operator Performance Under Stress

www.mit.ucf.edu

Operator Performance Under Stress

Current Progress and Future Prospects

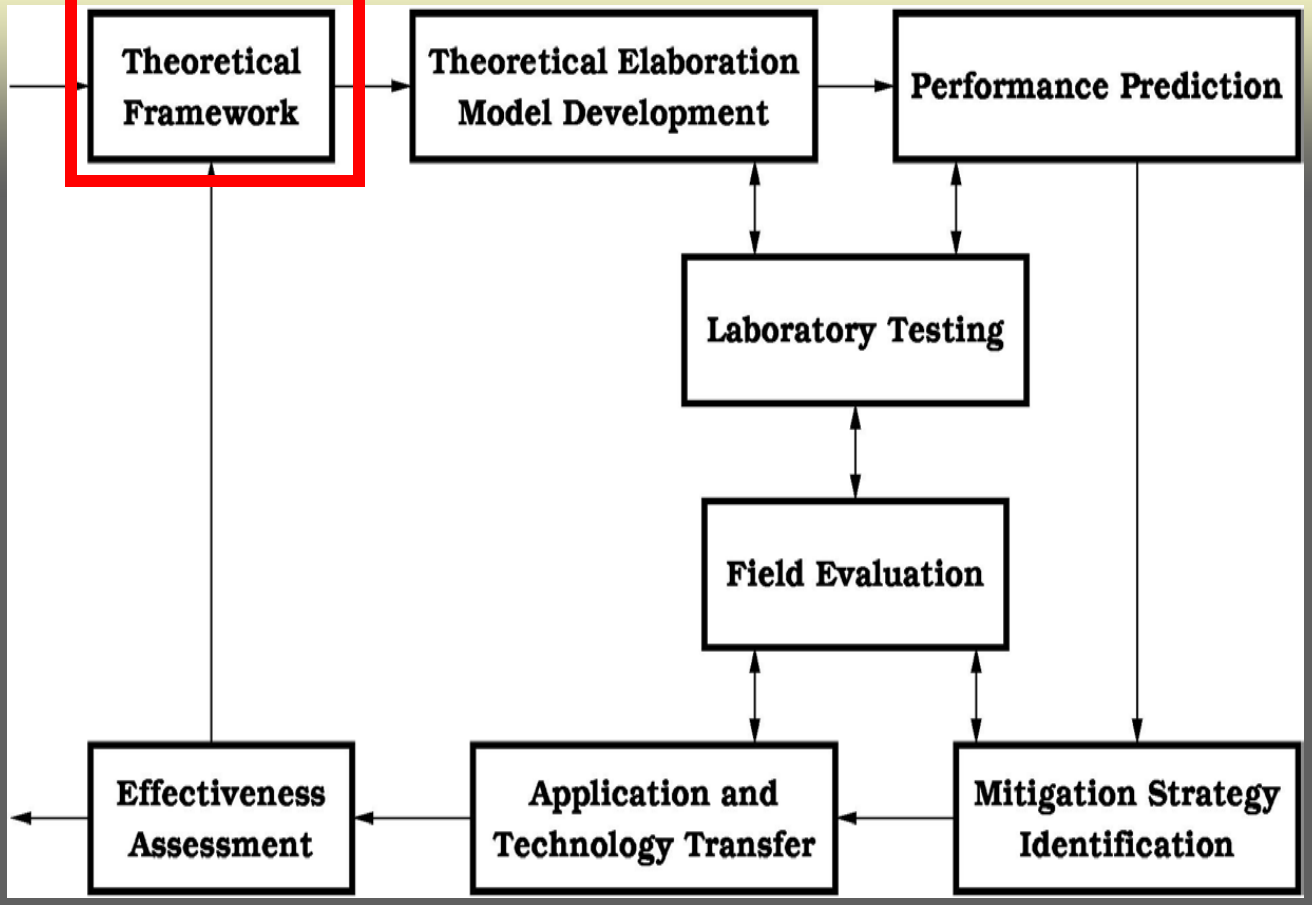
Peter A Hancock, Ph.D.

University of Central Florida



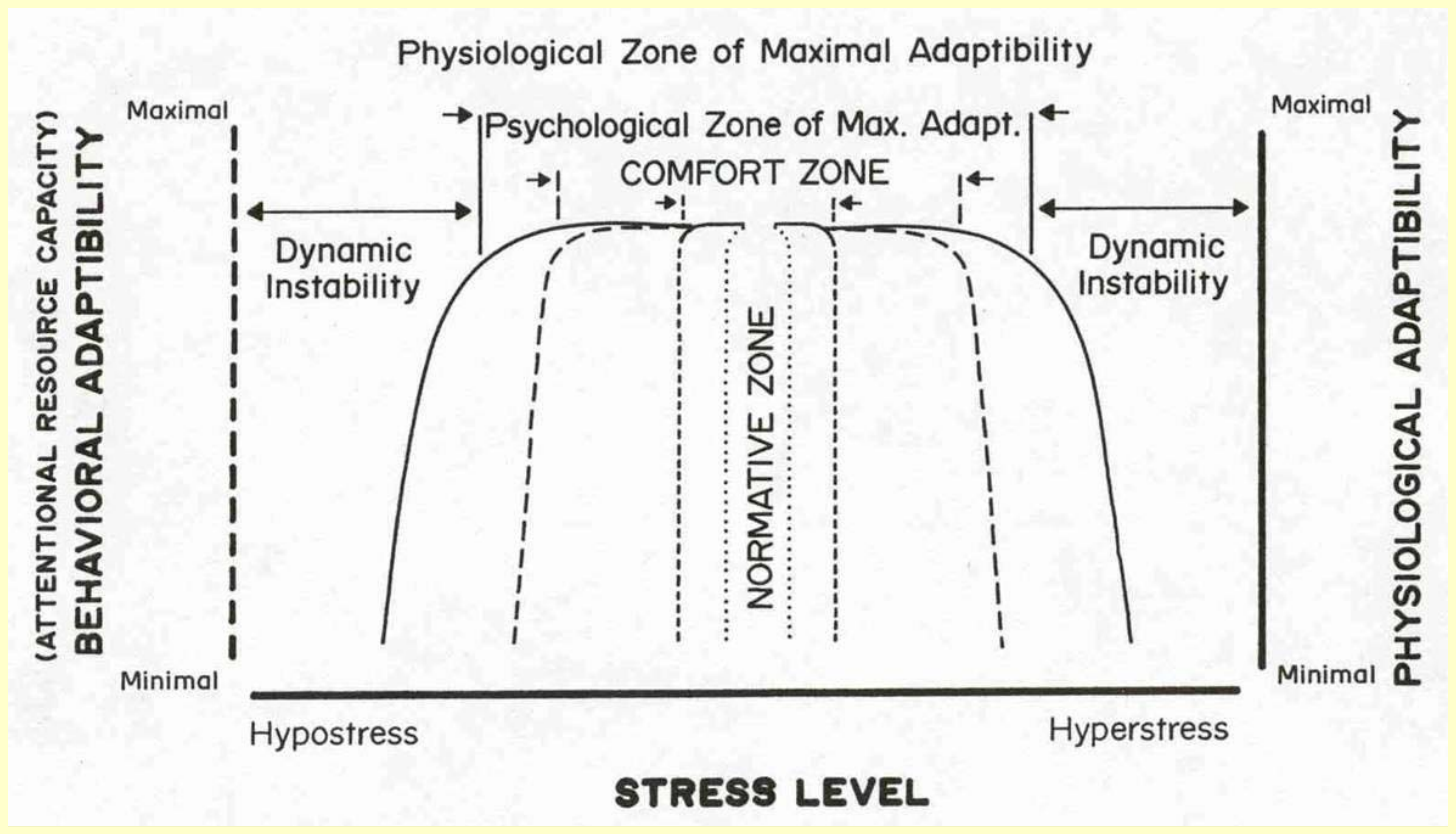
P.O.Box 161390 * Orlando, FL 32816 * 407-823-2310 * www.mit.ucf.edu

A Programmatic Process for Stress Research

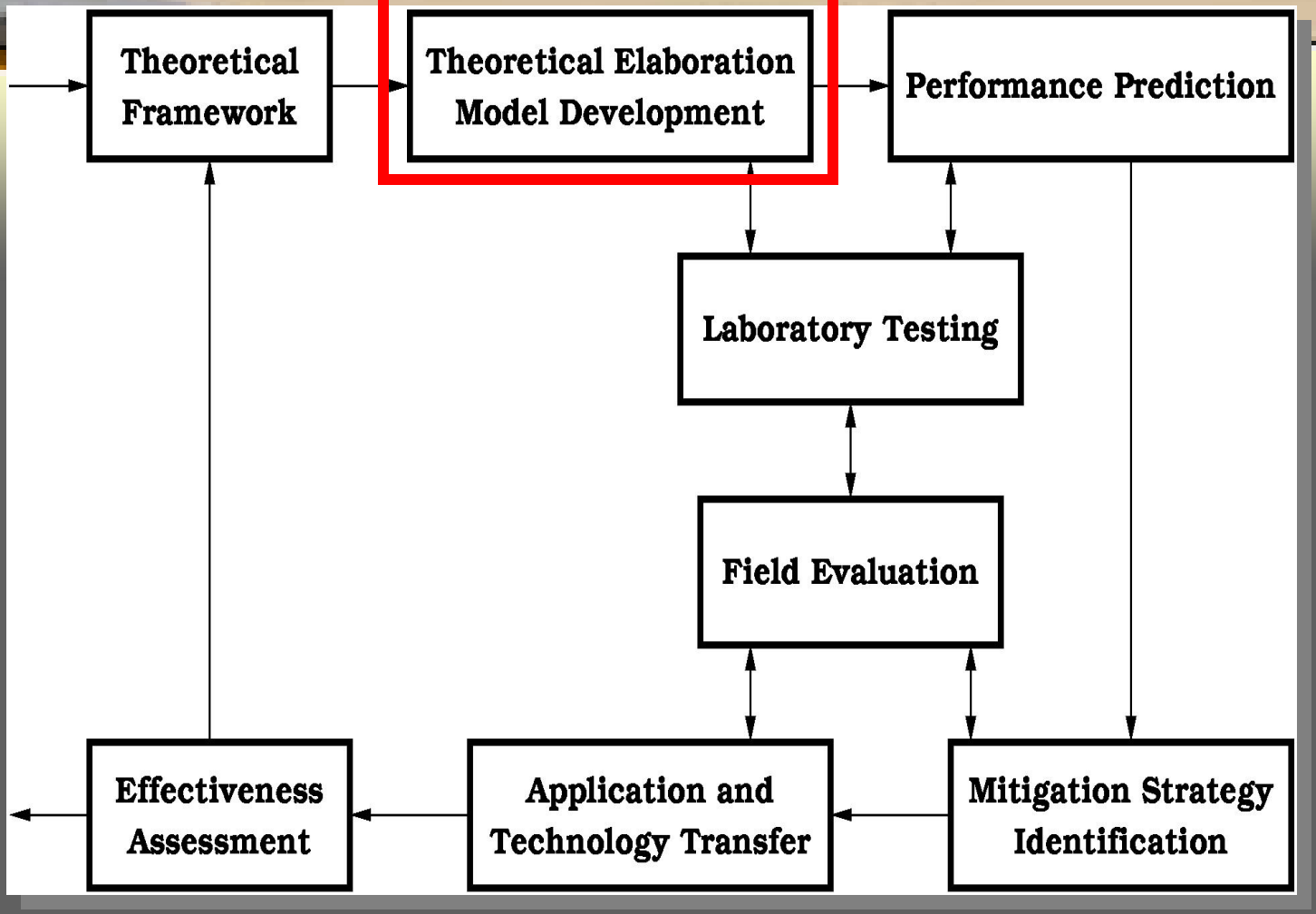


From: MURI-OPUS Research Program Overview

Theoretical Framework

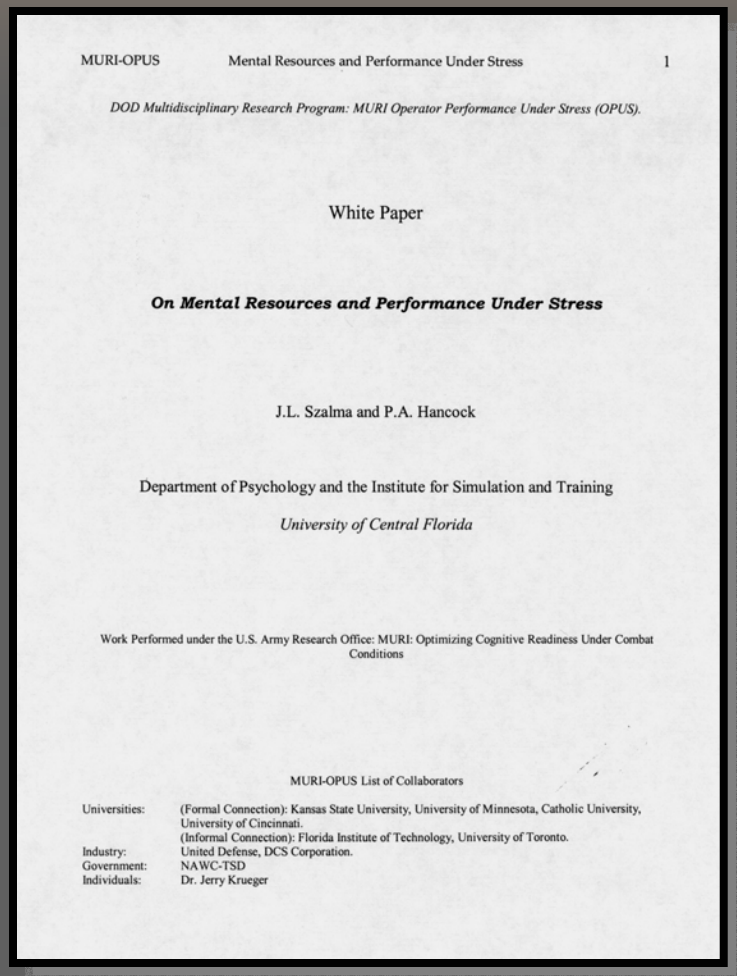


Hancock, P. A., & Warm, J. S. (1989) A dynamic model of stress and sustained attention. *Human Factors*, **31**(5), 519-537.



From: MURI-OPUS Research Program Overview

The White Paper Series



MURI-OPUS Mental Resources and Performance Under Stress 1

DOD Multidisciplinary Research Program: MURI Operator Performance Under Stress (OPUS).

White Paper

On Mental Resources and Performance Under Stress

J.L. Szalma and P.A. Hancock

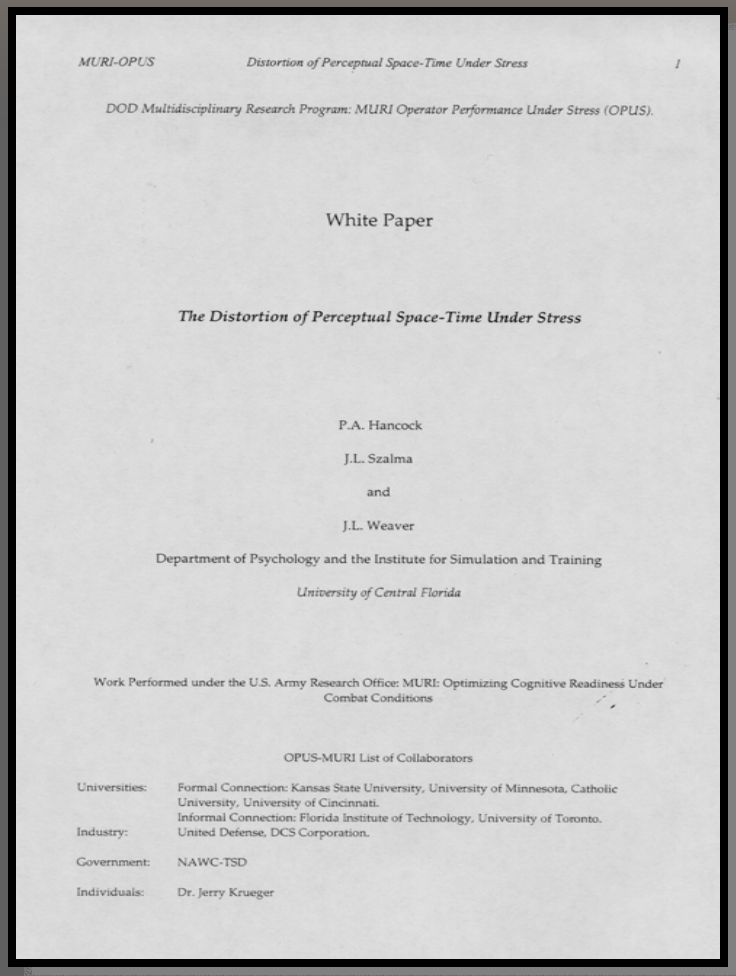
Department of Psychology and the Institute for Simulation and Training

University of Central Florida

Work Performed under the U.S. Army Research Office: MURI: Optimizing Cognitive Readiness Under Combat Conditions

MURI-OPUS List of Collaborators

Universities:	(Formal Connection): Kansas State University, University of Minnesota, Catholic University, University of Cincinnati. (Informal Connection): Florida Institute of Technology, University of Toronto.
Industry:	United Defense, DCS Corporation.
Government:	NAWC-TSD
Individuals:	Dr. Jerry Krueger



MURI-OPUS Distortion of Perceptual Space-Time Under Stress 1

DOD Multidisciplinary Research Program: MURI Operator Performance Under Stress (OPUS).

White Paper

The Distortion of Perceptual Space-Time Under Stress

P.A. Hancock

J.L. Szalma

and

J.L. Weaver

Department of Psychology and the Institute for Simulation and Training

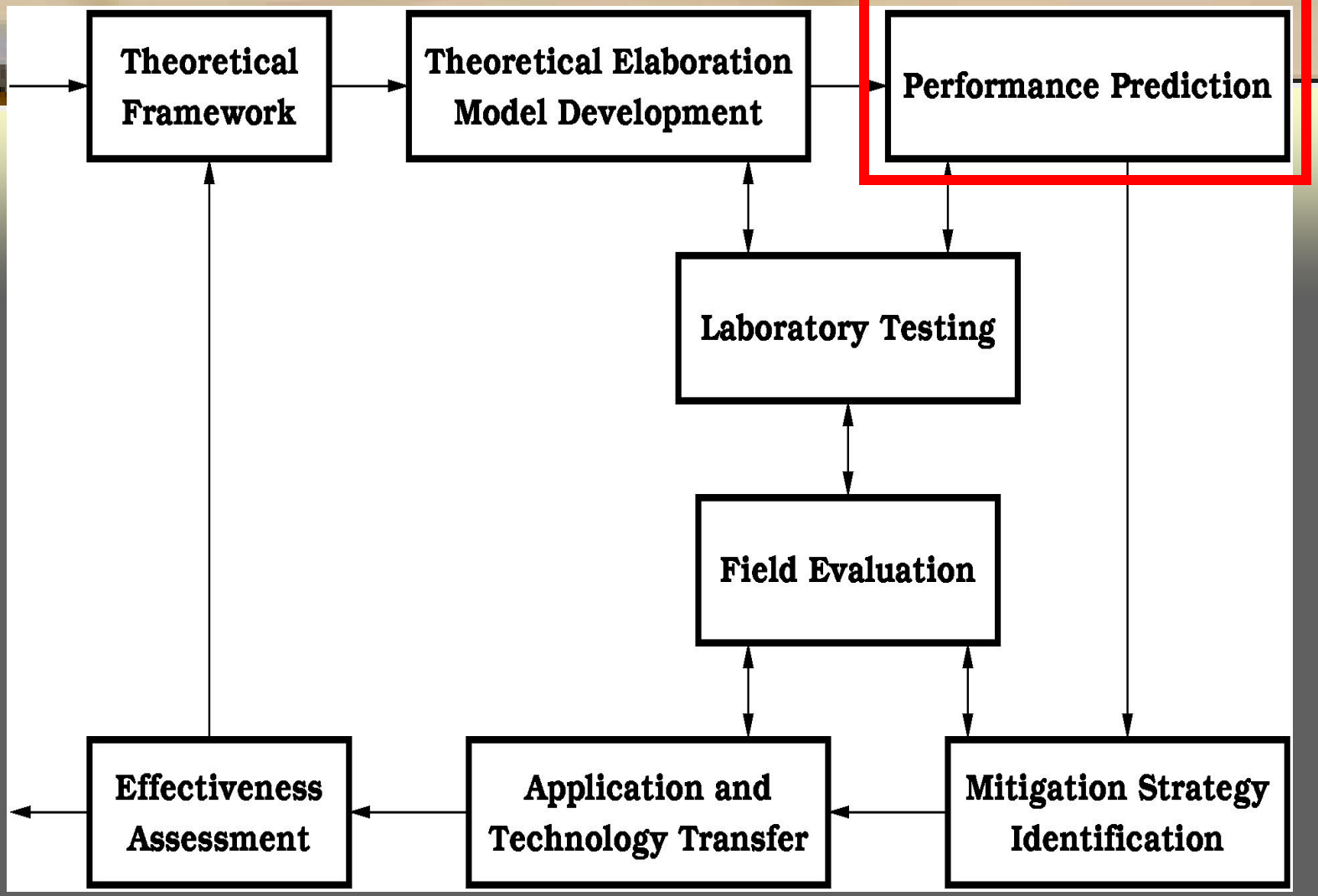
University of Central Florida

Work Performed under the U.S. Army Research Office: MURI: Optimizing Cognitive Readiness Under Combat Conditions

OPUS-MURI List of Collaborators

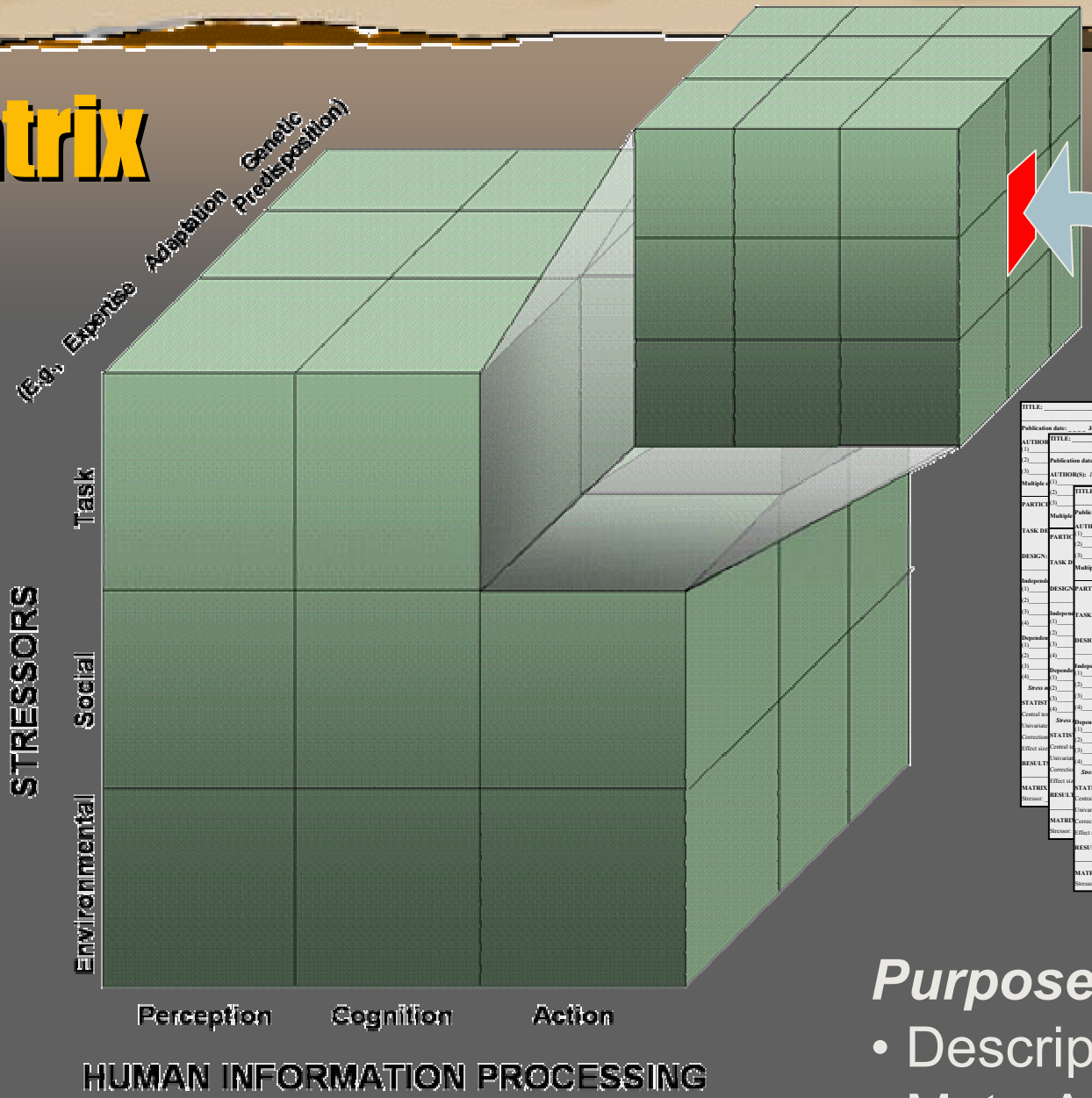
Universities:	Formal Connection: Kansas State University, University of Minnesota, Catholic University, University of Cincinnati. Informal Connection: Florida Institute of Technology, University of Toronto.
Industry:	United Defense, DCS Corporation.
Government:	NAWC-TSD
Individuals:	Dr. Jerry Krueger

Three of Five Complete, Two in Final Draft Phase



From: MURI-OPUS Research Program Overview

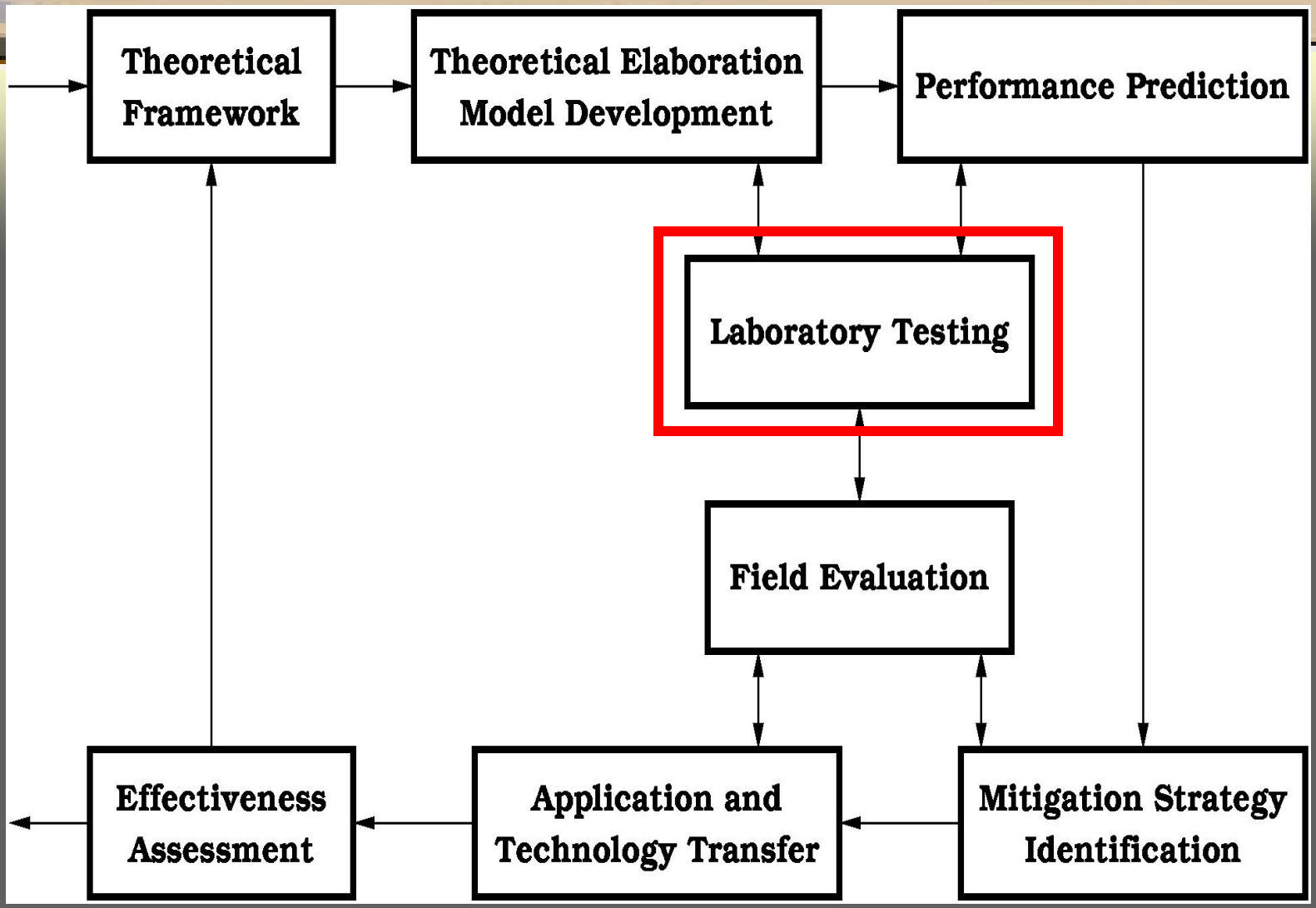
The Matrix



Purpose:

- Description
- Meta-Analysis

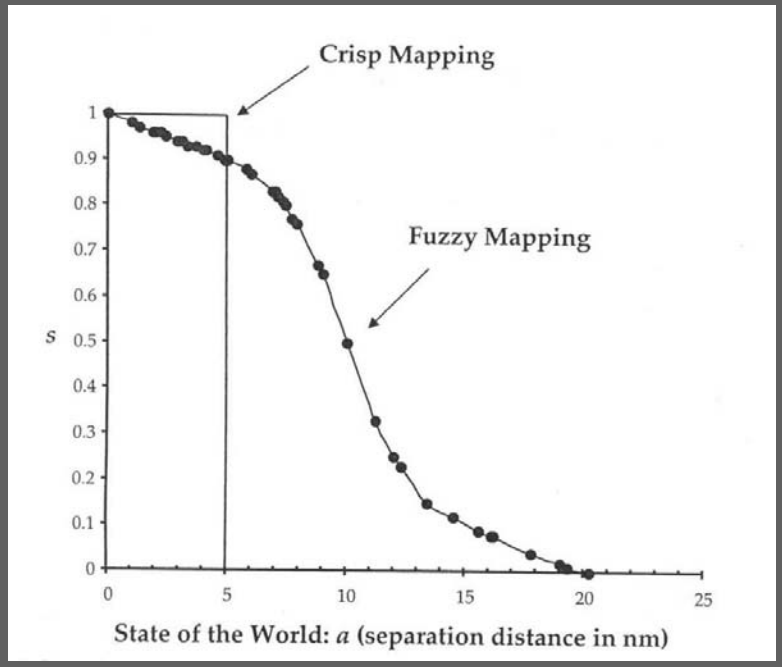
Laboratory Testing



Fuzzy Signal Detection Theory

Paper Presented Here by Lauren Murphy – MURI-OPUS Group





- Tests of the Fuzzy Signal Detection Theory model: Vigilance
- Comparison of the FSDT model to the traditional SDT model

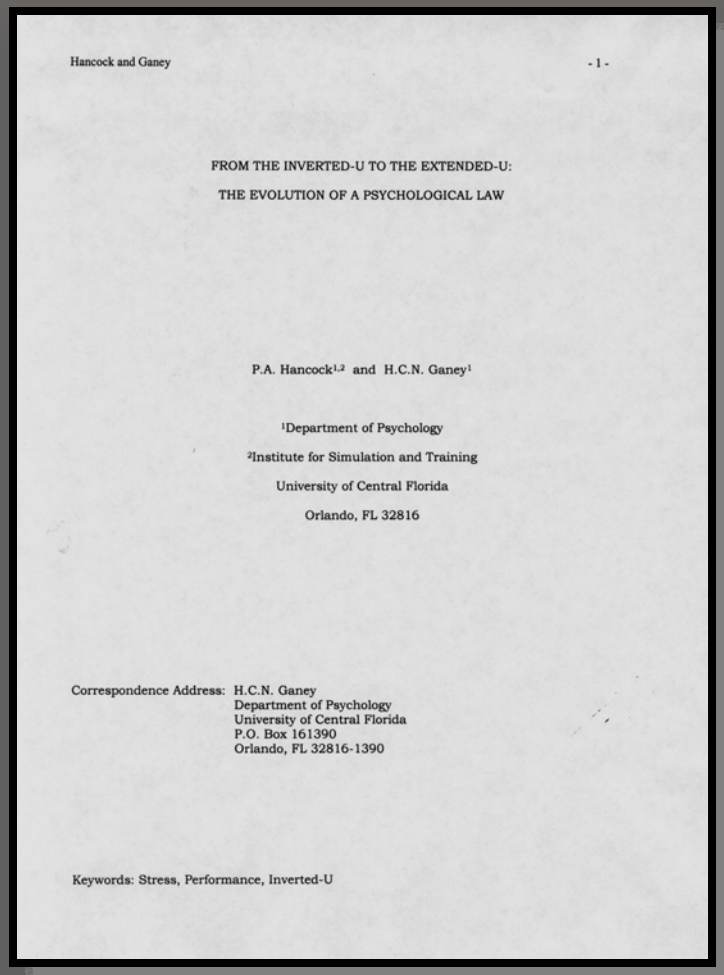
SKR Performance Variation under Impulse Noise

Paper Presented Here by Kelly Burke – MURI-OPUS Group

- Investigated the degree that aspects of complex task performance, corresponding to Rasmussen's skills, rules, and knowledge typology, would degrade under conditions of stress (noise).
- It was hypothesized that the resource management task classified as knowledge-based will be shed first in response to noise, followed by the monitoring task classified as rule-based, and the skill-based tracking task will be the last to be shed.
- The performance findings did not support our original degradation hypotheses.
- Serves to illustrate that stressor effects are often not intuitive and in fact often counter to what is hypothesized. Most important is to consider the utility of the MATB as truly illustrative of the components of Rasmussen's typology.



Inverted-U to the Extended-U



Hancock and Ganey - 1 -

FROM THE INVERTED-U TO THE EXTENDED-U:
THE EVOLUTION OF A PSYCHOLOGICAL LAW

P.A. Hancock[1,2] and H.C.N. Ganey[1]

[1]Department of Psychology
[2]Institute for Simulation and Training
University of Central Florida
Orlando, FL 32816

Correspondence Address: H.C.N. Ganey
Department of Psychology
University of Central Florida
P.O. Box 161390
Orlando, FL 32816-1390

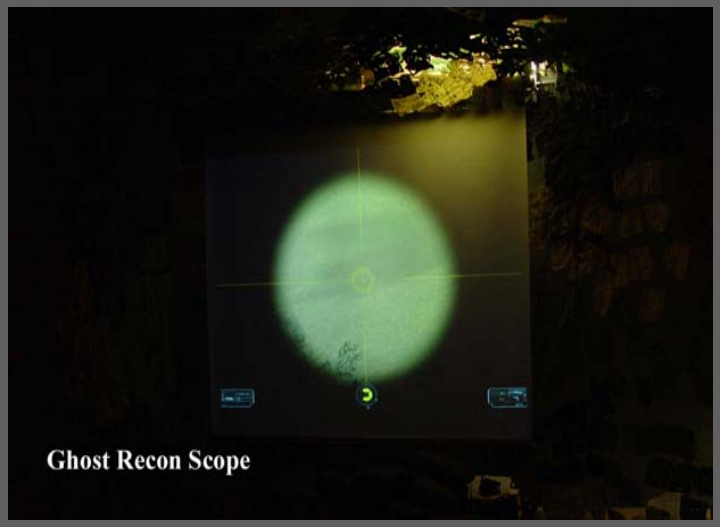
Keywords: Stress, Performance, Inverted-U

Hancock, P.A. & Ganey, H.C.N. (2003) From the inverted-U to the extended-U: The evolution of a psychological law.

Human Performance in Extreme Environments, in press.

Ghost Recon Laboratory



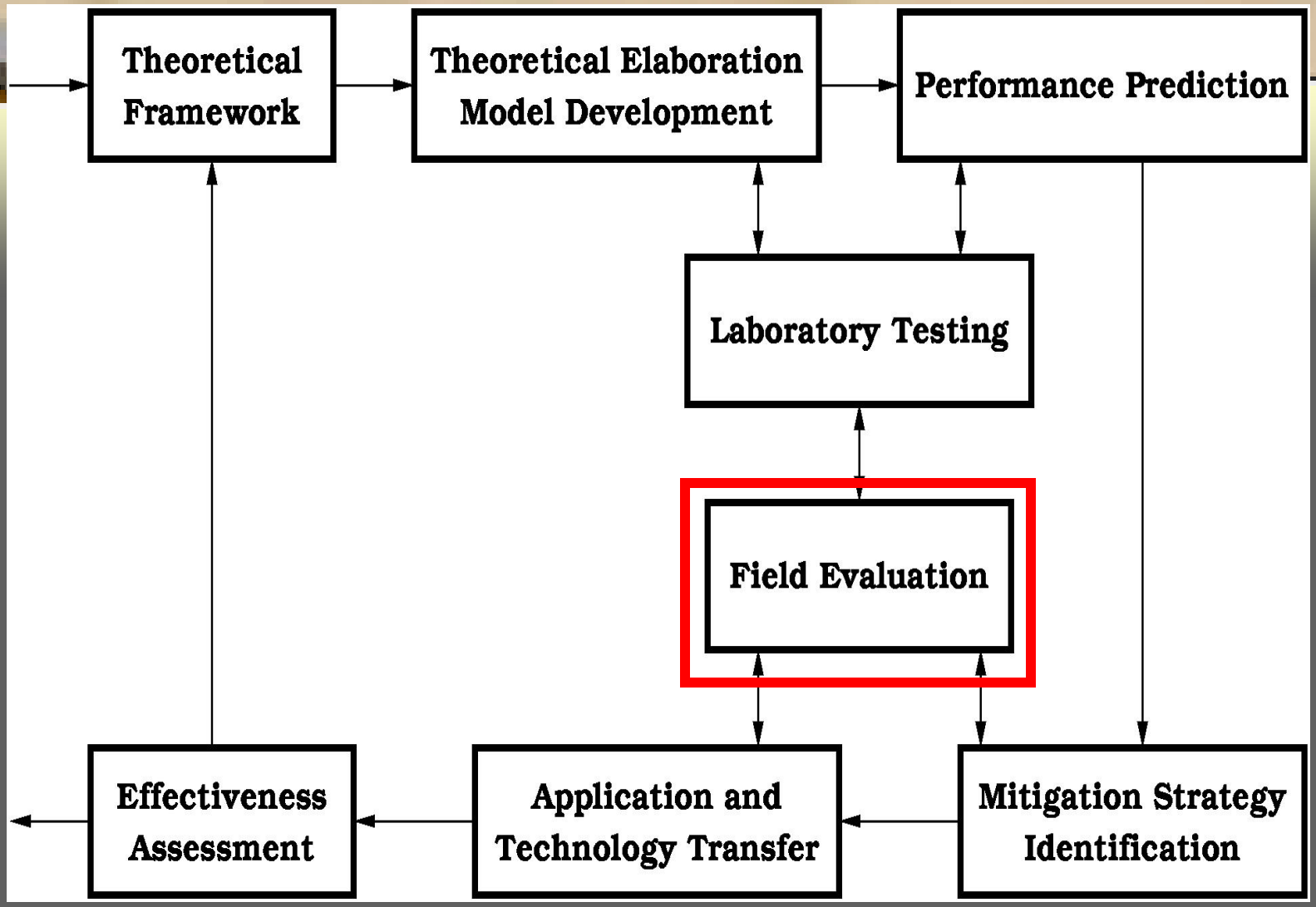


- Medium Fidelity Soldier Simulator
- Platform: Ghost Recon
- Study: GR-NAV (virtual map)

  Target discrimination and engagement

  GR-Decision Making

  GR-Time Stress


Theoretical Framework
Theoretical Elaboration Model Development
Performance Prediction
Laboratory Testing
Field Evaluation
Effectiveness Assessment
Application and Technology Transfer
Mitigation Strategy Identification

Field Evaluations

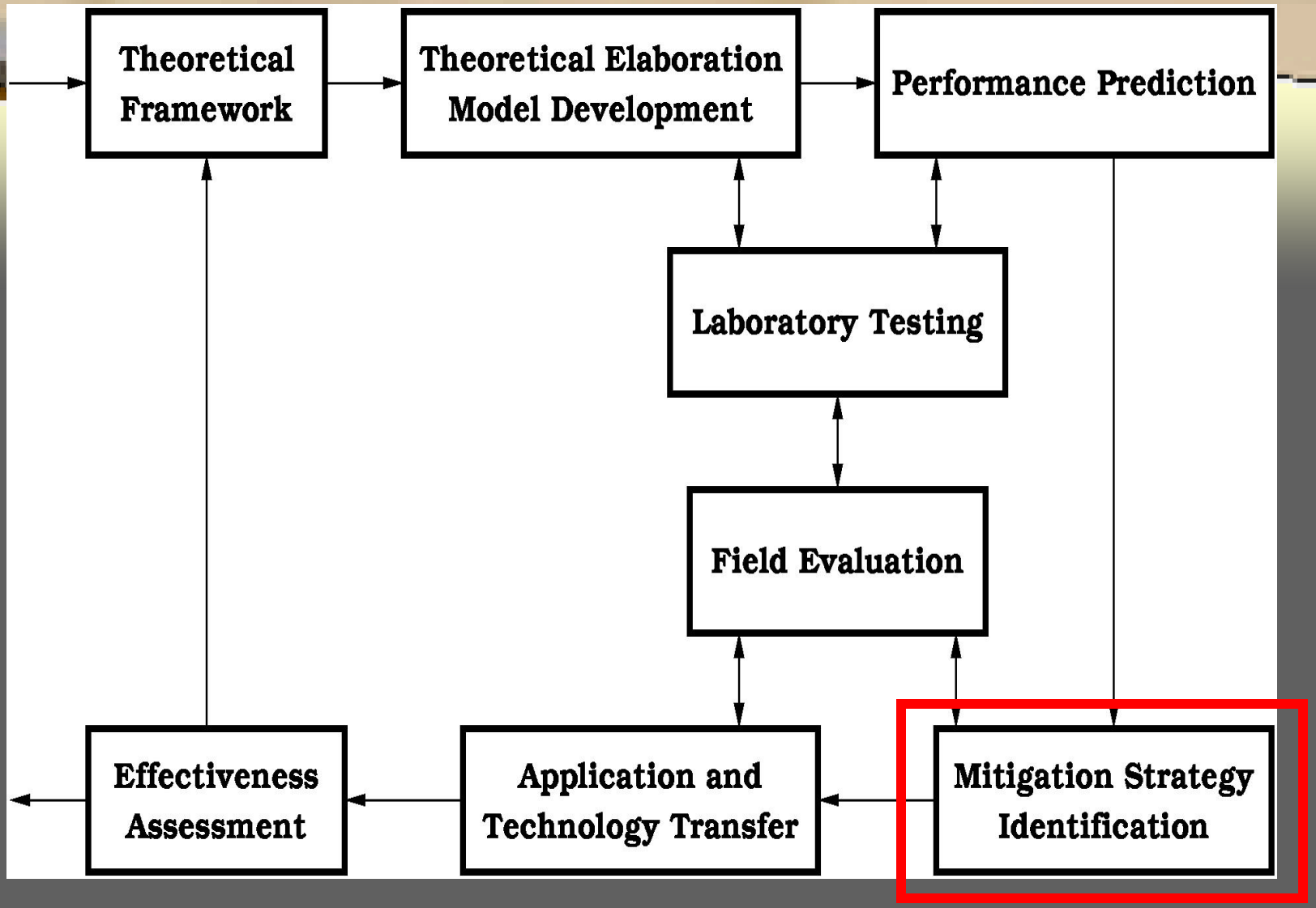
- Ft. Polk Exercise - Dr. Wayne Harris







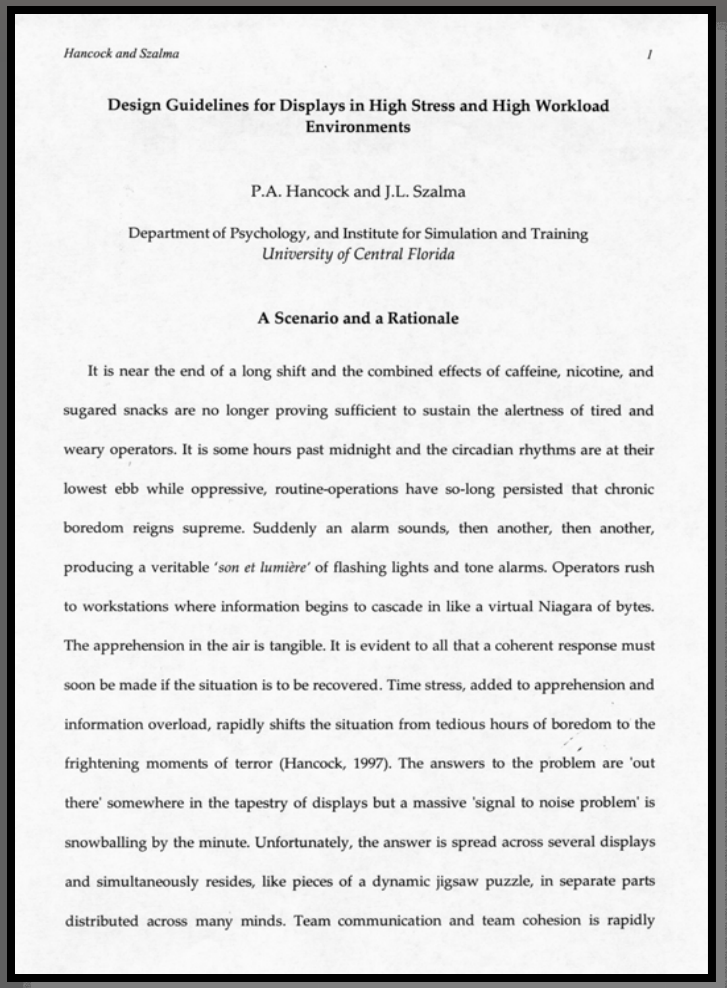
From Field to Theory


Theoretical Framework
Theoretical Elaboration Model Development
Performance Prediction
Laboratory Testing
Field Evaluation
Effectiveness Assessment
Application and Technology Transfer
Mitigation Strategy Identification

Mitigation Strategies: An Example

- Stress/Performance Design Guidelines





Hancock and Szalma 1

Design Guidelines for Displays in High Stress and High Workload Environments

P.A. Hancock and J.L. Szalma

Department of Psychology, and Institute for Simulation and Training
University of Central Florida

A Scenario and a Rationale

It is near the end of a long shift and the combined effects of caffeine, nicotine, and sugared snacks are no longer proving sufficient to sustain the alertness of tired and weary operators. It is some hours past midnight and the circadian rhythms are at their lowest ebb while oppressive, routine-operations have so-long persisted that chronic boredom reigns supreme. Suddenly an alarm sounds, then another, then another, producing a veritable '*son et lumière*' of flashing lights and tone alarms. Operators rush to workstations where information begins to cascade in like a virtual Niagara of bytes. The apprehension in the air is tangible. It is evident to all that a coherent response must soon be made if the situation is to be recovered. Time stress, added to apprehension and information overload, rapidly shifts the situation from tedious hours of boredom to the frightening moments of terror (Hancock, 1997). The answers to the problem are 'out there' somewhere in the tapestry of displays but a massive 'signal to noise problem' is snowballing by the minute. Unfortunately, the answer is spread across several displays and simultaneously resides, like pieces of a dynamic jigsaw puzzle, in separate parts distributed across many minds. Team communication and team cohesion is rapidly

Hancock, P.A. & Szalma, J.L. (2003). Design guidelines for displays in high stress and high workload environments. Ergonomics and Design. (in press).

Current State and Near-Term Goals

- SAST-Based Research (Information Stress).
- Integration of Theoretical Advances.
- Implementation of Mitigation Process.
- Elaborated Empirical Investigation.
- Stress Meta-Analysis
- Inputs to Advanced Models
- Improved Performance Prediction Capacity

ACKNOWLEDGEMENT

This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI), Operator Performance Under Stress (OPUS) program; P.A. Hancock, *Principal Investigator*, administered by the Army Research Office under Grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, Dr. Mike Drillings and Dr. Paul Gade for providing administrative and technical direction for the Grant.

Predicting performance under stressful conditions: A theoretical and empirical approach

Paul Ward

James L. Szalma

Peter A. Hancock

Multi-Disciplinary University Research Initiative
MURI-OPUS
Operator Performance Under Stress



Approaches to Stress Research

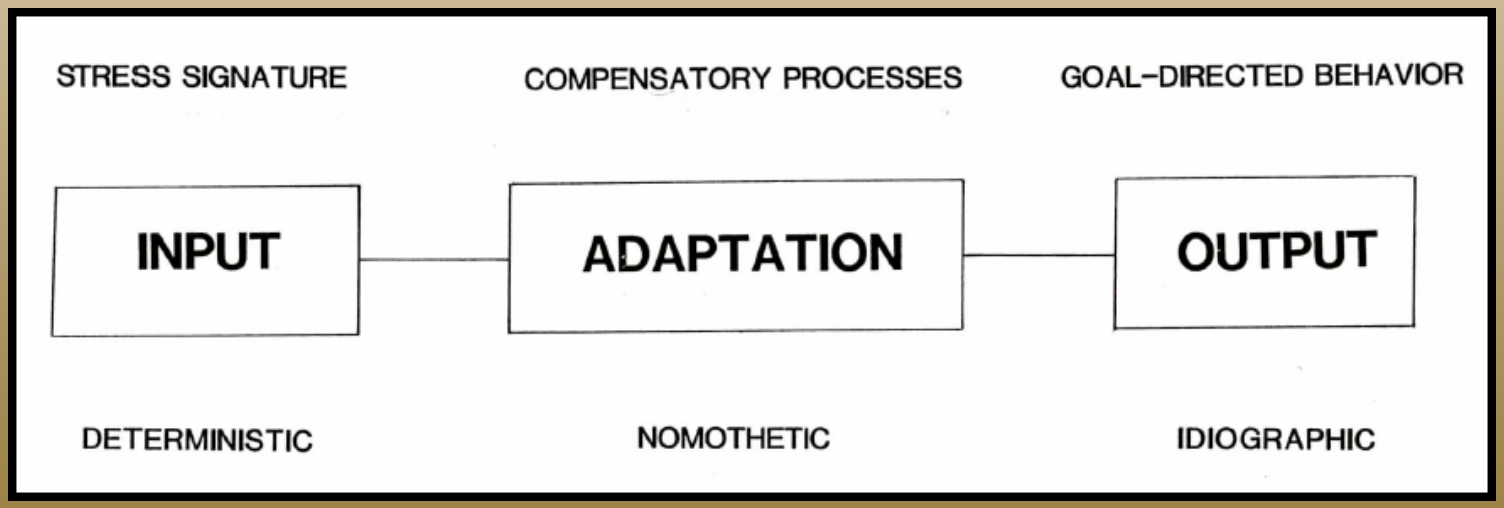
- Stress:
  - Property of the environment
  - Appraisal / coping mechanisms
  - Physiological response

- Multiple levels of analysis of a single dynamic process (Hancock & Warm, 1989)



The Maximal Adaptability Model:

The Trinity of Stress

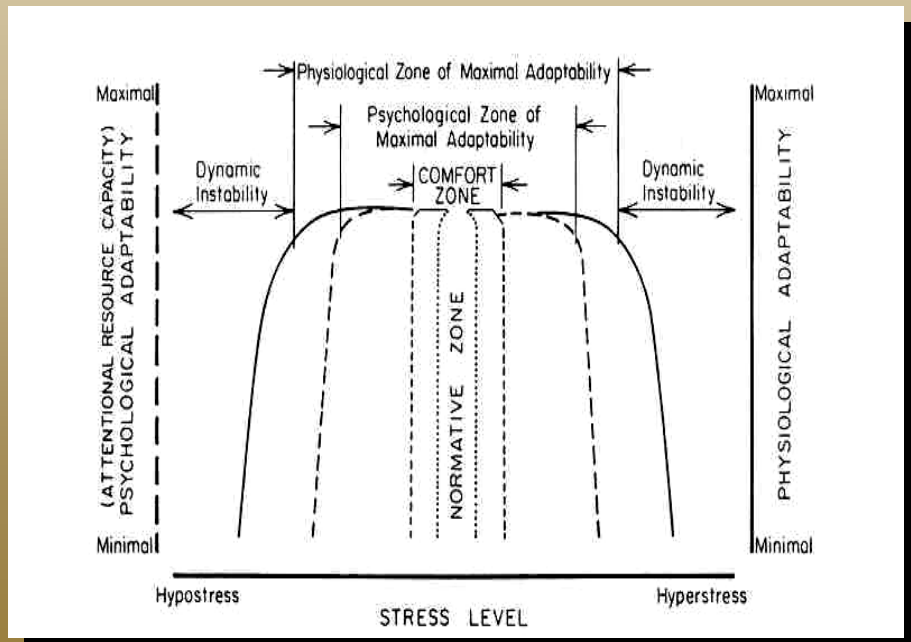


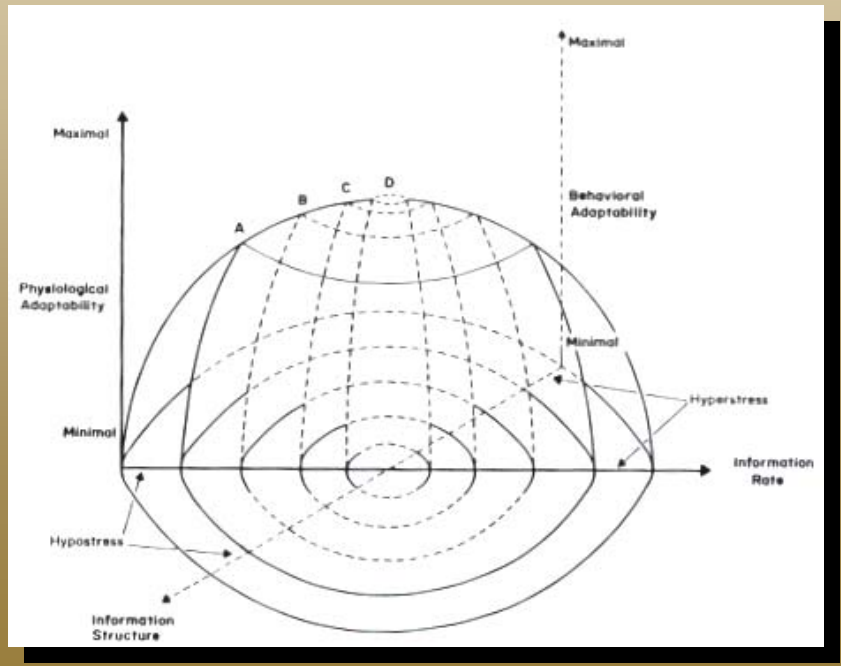
Hancock, P.A., & Warm, J.S., (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.



The Maximal Adaptability Model:

The Extended-U





Hancock, P.A., & Warm, J.S., (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.



Effects of Stress on Task Performance

- Multiple task solutions are available in low stress, information-rich environments
- Increased stress reduces the available strategies and number of solution paths to meet the goal
- Stereotypical behavior increases as attention narrows to find most appropriate and available task solution
- Individuals use compensatory strategies to maintain optimum information flow

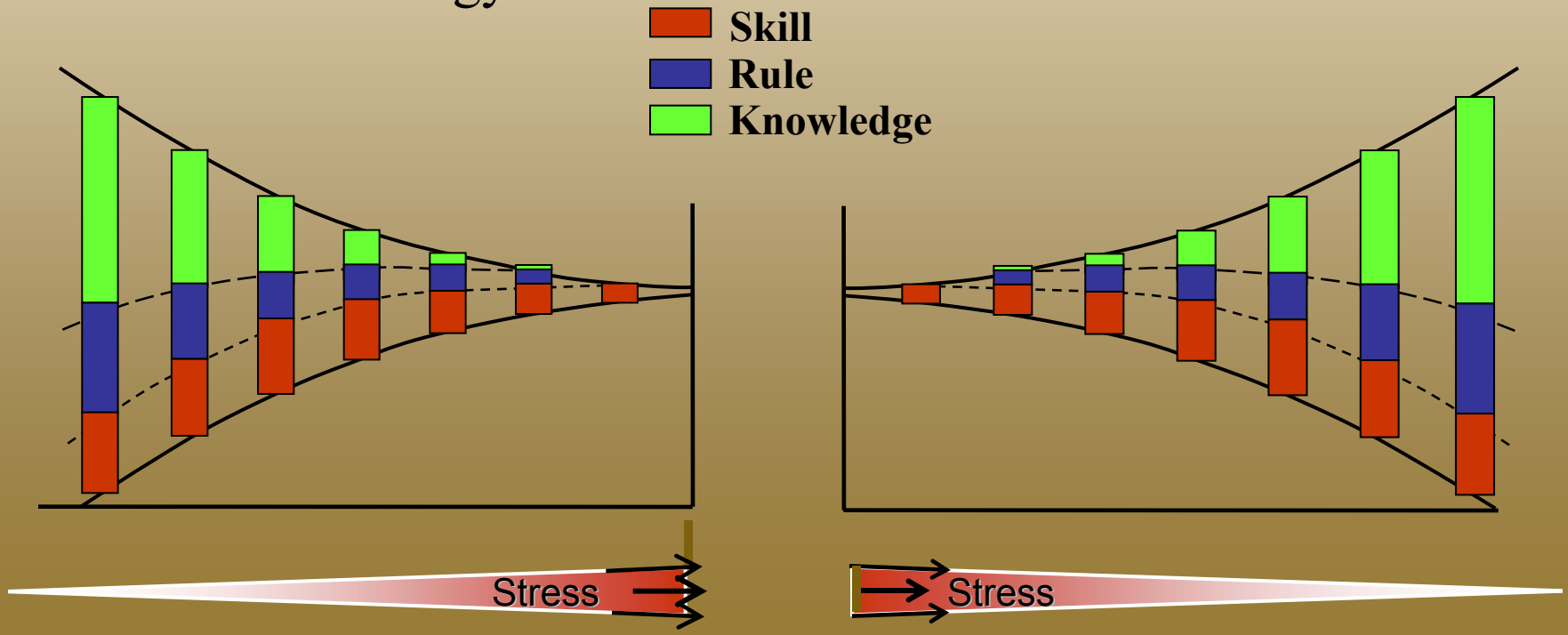
  e.g., modification of information structure and rate



Predicted Effects

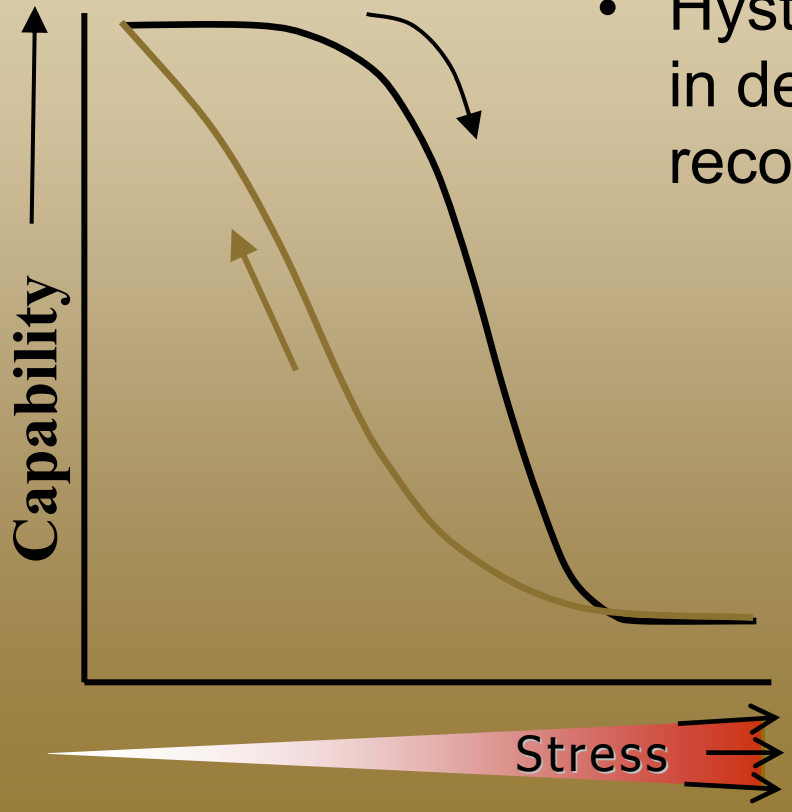
Load Shedding and Attentional Strategy

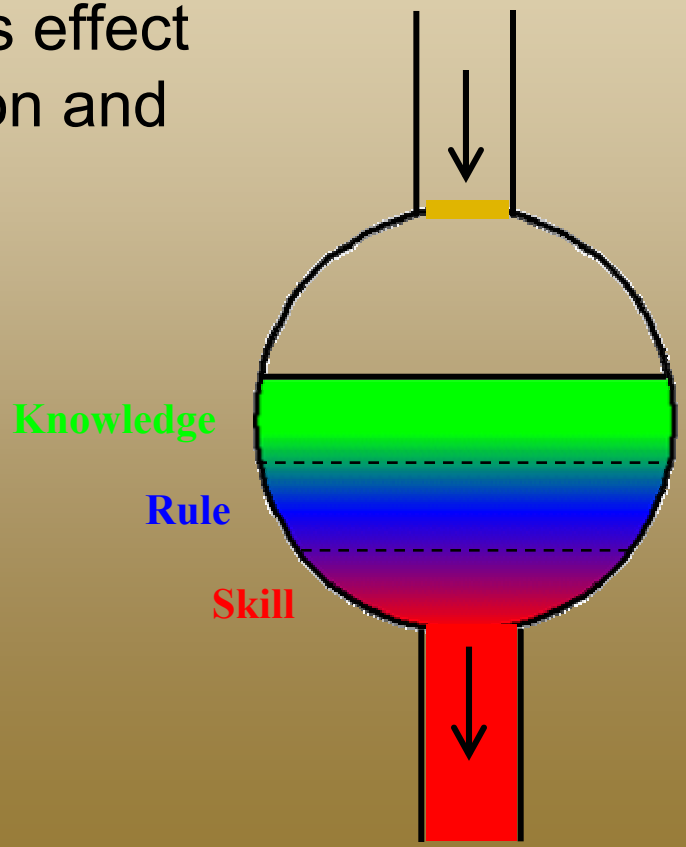
Recovery Strategy



Predicted Effects

- Hysteresis effect in depletion and recovery?







Outstanding issues

Cognitive patterns of stress states

Stressor or condition	Brief description of manipulation	Alertness	Selectivity	Fast responding Speed	Fast responding Accuracy	STM
Anxiety	High scorers on trait-state anxiety tests; induced by danger/threat	+	+	0	−	−
Noise	Continuous broad-band noise at about 80-100 dB	+	+	0	−	−
Amphetamine	Oral ingestion in normal clinical dosages	+	+	+	(?)	−
Nicotine	Smoking one or two cigarettes, or oral ingestion of nicotine tablets	+	+	+	0(?)	−
Incentives	Augmented feedback, financial reward, or induced competitiveness	+	+	+	+	+
Time of day (p.m.)	Testing at later times of day (between about 1500 and 2100)	+	(?)	+	−	−
Heat	WGBT of about 30-40 °C for up to 2-3 hours	+(?)	+	0	−	0
Alcohol	Oral ingestion of about 0.2-0.5 g/kg (peak BA level about 0.04%)	−	+(?)	−	−	−
Minor tranquilizers	Oral ingestion of benzodiazepines in normal clinical dosages	−	−	−	−	−
Sleep loss / night-work	One night without sleep / working at night before adaptation	−	−	−	−	0
Prolonged work / boredom	Continuous work for 30+ minutes on repetitive tasks	−	+(?)	−	−	0
Time of day (a.m.)	Testing at earlier times of day (between	−	(?)	−	+	+

Hockey, G.R., & Hamilton, P.A. (1983). The cognitive patterning of stress states. In G.R. Hockey (Ed.), *Stress and fatigue in human performance,* 331-361. New York: Wiley.



Outstanding issues

- Transactional viewpoint (Lazarus & Folkman, 1984; Matthews, 2001)
  - Person-environment interaction
  - Levels of transaction (e.g., bio-cognitive, cognitive-adaptive)
- What factors change the nature of the stress-performance relationship?
  - Internal Moderators: traits, emotions, coping style, intelligence, etc.

    e.g., Anxiety – attentional interference, cognitive bias, motivational role
  - External Moderators: training, expertise, skill-level, etc.

    e.g., Do predicted effects of anxiety on WM apply to expert adaptation to the task (i.e., LTWM)?



The Big Picture: Performance Prediction and Procurement

- Outstanding questions:
  - By what degree does stress affect performance?
  - What can be done to alleviate these effects?
  - Are some individuals more suited to a task than others?
  - Can we aid in the selection of the best man for the job?
  - Can we train an individual to acquire appropriate strategies/skills?
- IMPRINT (Improved Performance Research Integration Tool)
  - A task network modeling architecture that aims to address some of these issues



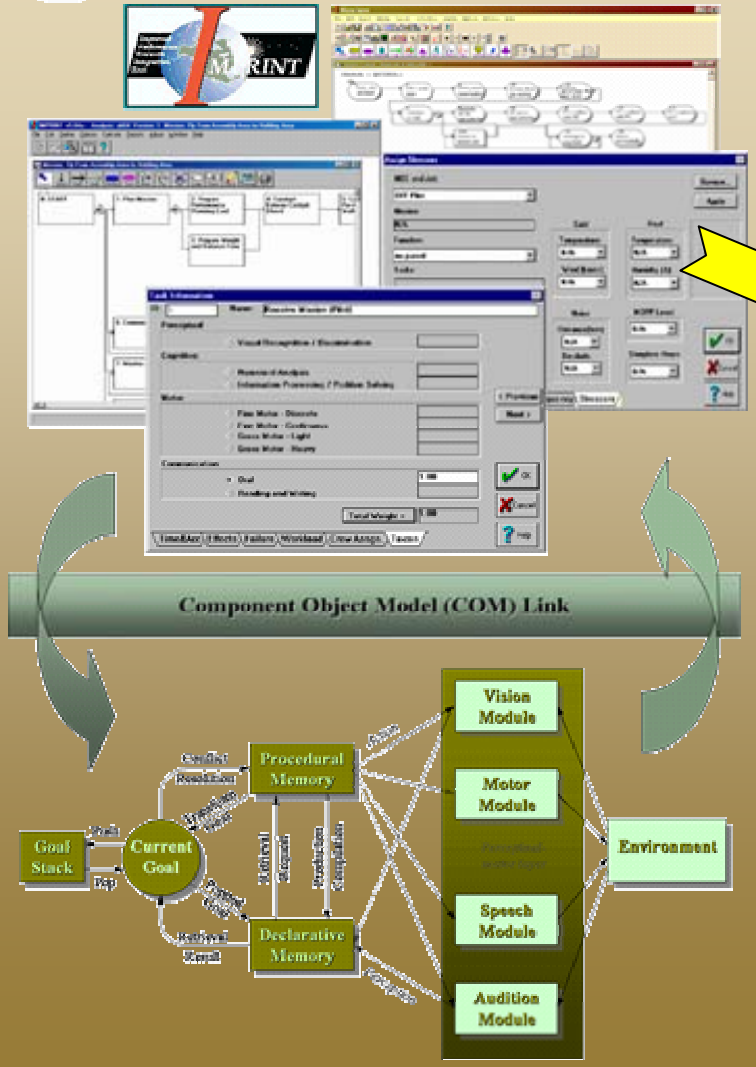
MIT²

INSTITUTE for SIMULATION & TRAINING

Performance Shaping Functions

Stress Degradation Algorithms

Workload 'look-up tables'





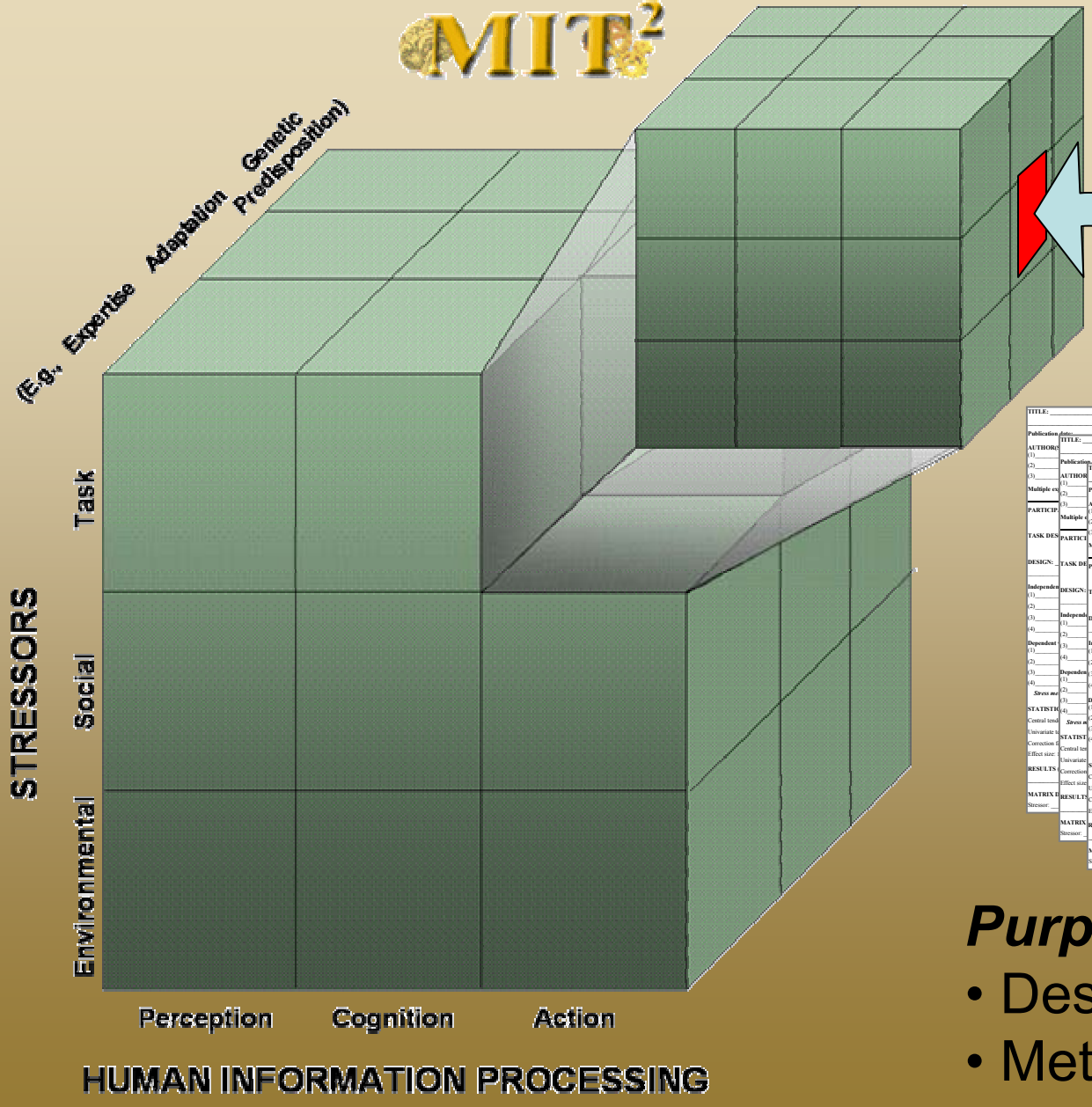
Multi-Disciplinary University Research Initiative

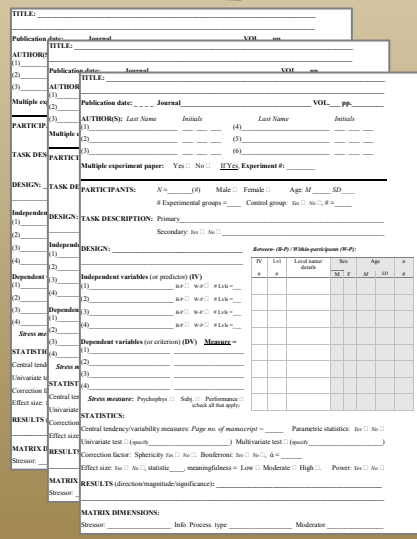
MURI-OPUS

Operator Performance Under Stress



The Matrix





Purpose:

- Description
- Meta-analysis

Multi-Disciplinary University Research Initiative

MURI-OPUS

Operator Performance Under Stress



Goals of the research

- Comprehensive theoretical and empirical assessment of stress and performance
- Provision of stress degradation algorithms (short-term) and address integration of stress/cognition via computation modeling (long-term)
- Increased prediction capacity of current and future models
- Development of suitable training strategies/guidelines
- Identification of selection criteria



ACKNOWLEDGEMENT

This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI), Operator Performance Under Stress (OPUS) program; P.A. Hancock, *Principal Investigator*, administered by the Army Research Office under Grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, Dr. Mike Drillings and Dr. Paul Gade for providing administrative and technical direction for the Grant.





Application of Fuzzy Signal Detection Theory to the Evaluation of Performance Under Stress

James L. Szalma

Paul Ward

Peter A. Hancock

University of Central Florida

Human Performance under Stress

- Information Overload/Complexity
- Physical/Social Environment
- Uncertainty



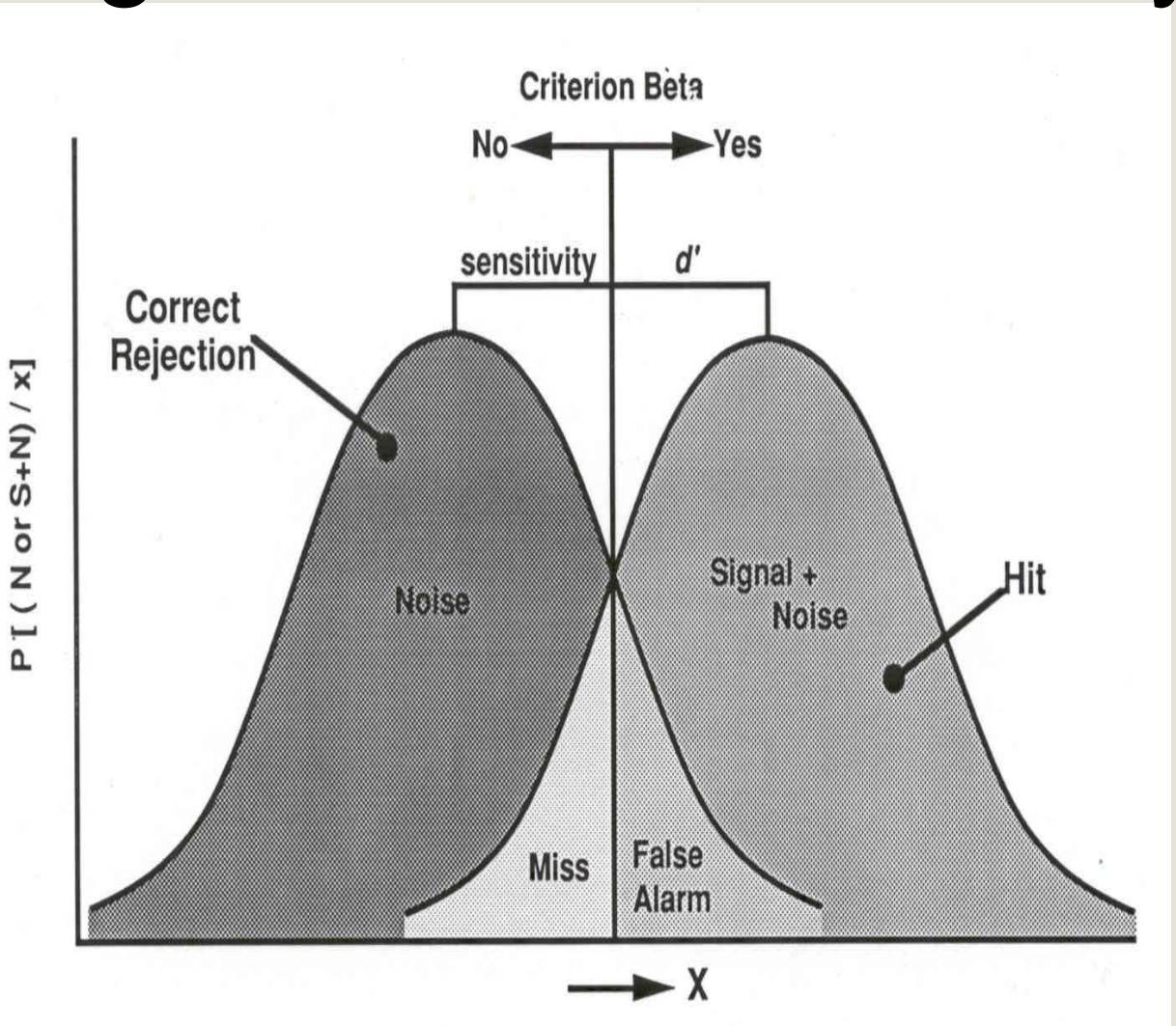




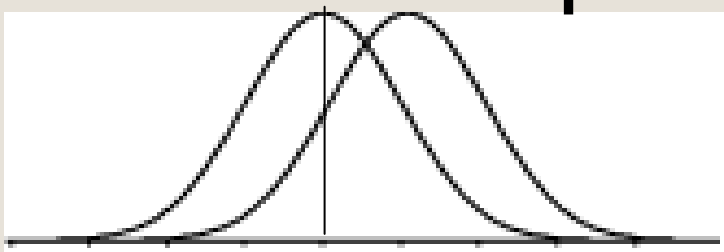
How Do We Assess Performance?

State of the World	Response: Yes	Response: No
Signal	Hit (H)	Miss (M)
Noise	False Alarm (FA)	Correct Rejection (CR)

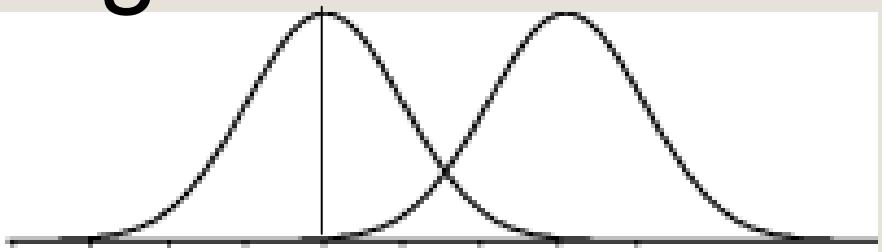
Signal Detection Theory



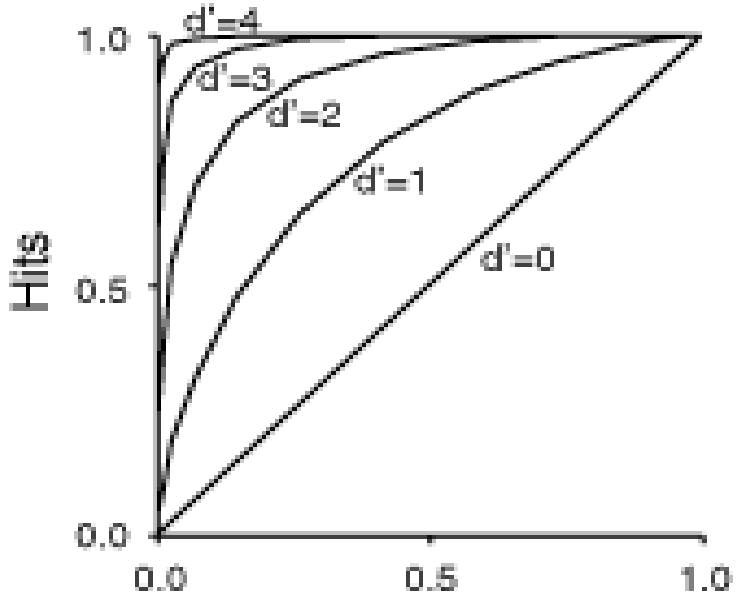
Receiver Operating Characteristic



d' = 1 (lots of overlap)



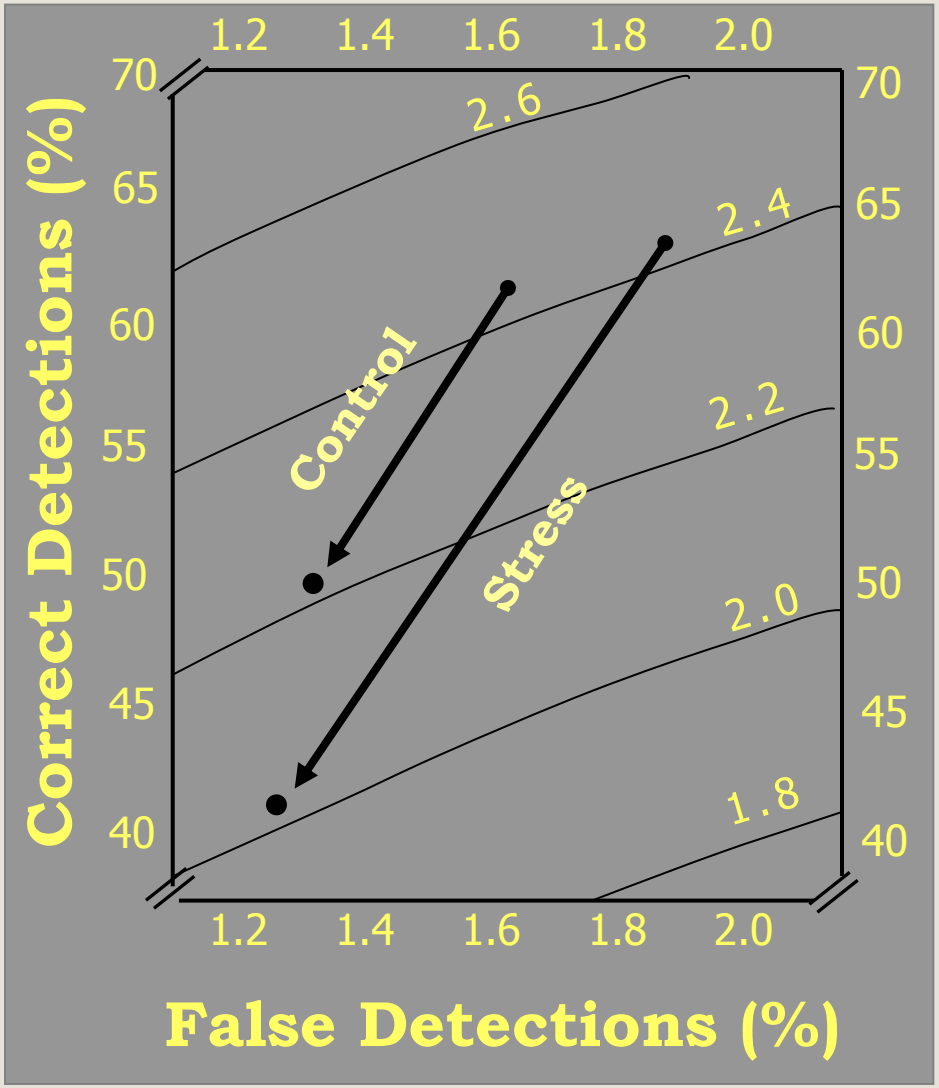
d' = 3 (not much overlap)



ROC curves

Signal Detection under Stress

- **How Does Stress Impact d'? and β?**
- **d' is reduced**
- **β increases (sometimes)**



Fuzzy Signal Detection Theory

Traditional Set Theory: $A \cap \bar{A} = 0$

Fuzzy Set Theory: $A \cap \bar{A} \neq 0$

Degrees of membership possible

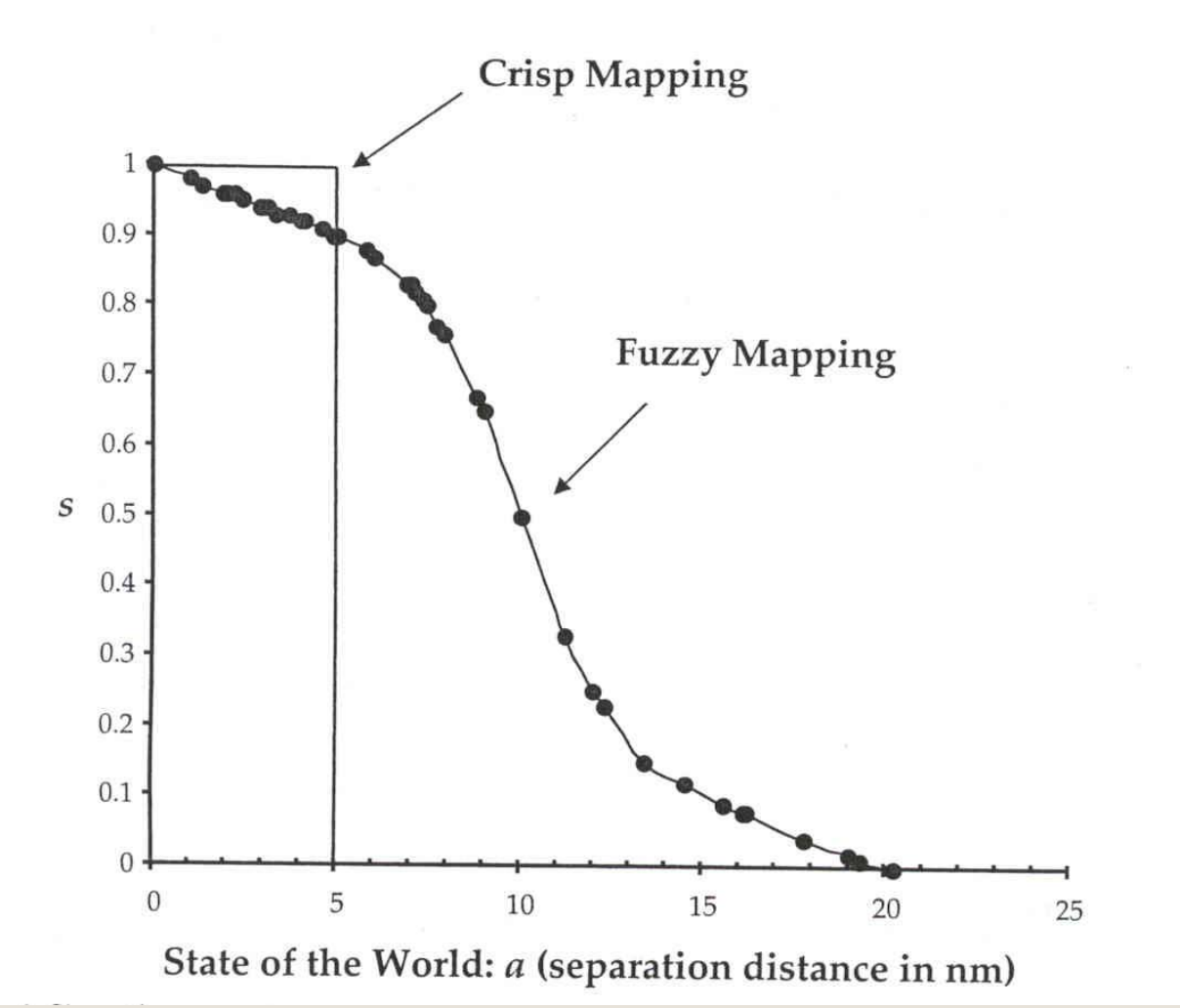
Elements of Fuzzy Signal Detection Theory

- Events can belong to the set “signal” (**s**) to a degree ranging from 0 to 1
- Events can belong to the set “response” (**r**) to a degree ranging from 0 to 1

Computation of FSDT Measures

- Select mapping functions for signal & response dimensions
- Assignment of degrees of membership to the four outcomes (H, M, FA, CR) using mixed implication functions.
- Compute fuzzy Hit, Miss, False Alarm, and Correct Rejection Rates
- Compute detection theory measures of sensitivity and response bias

Mapping Functions



Computation of Fuzzy Hit and False Alarm Rates

Mixed Implication Functions

H = min (**s**,**r**)
M = max (**s**-**r**, 0)
FA = max (**r**-**s**, 0)
CR = min (1-**s**, 1-**r**)

Current Empirical Work

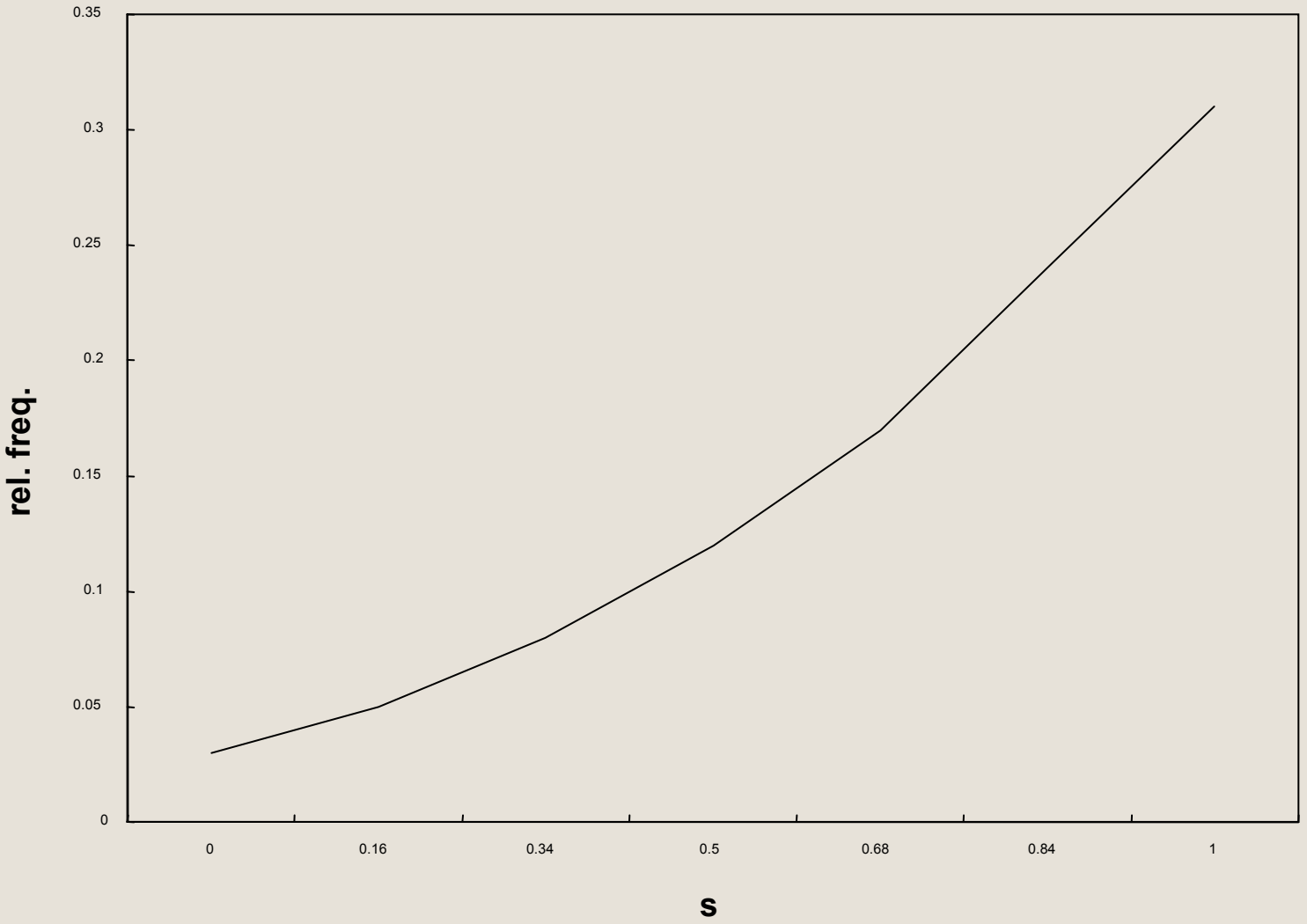
Construct Fuzzy ROC Curves

Compare to Crisp Model

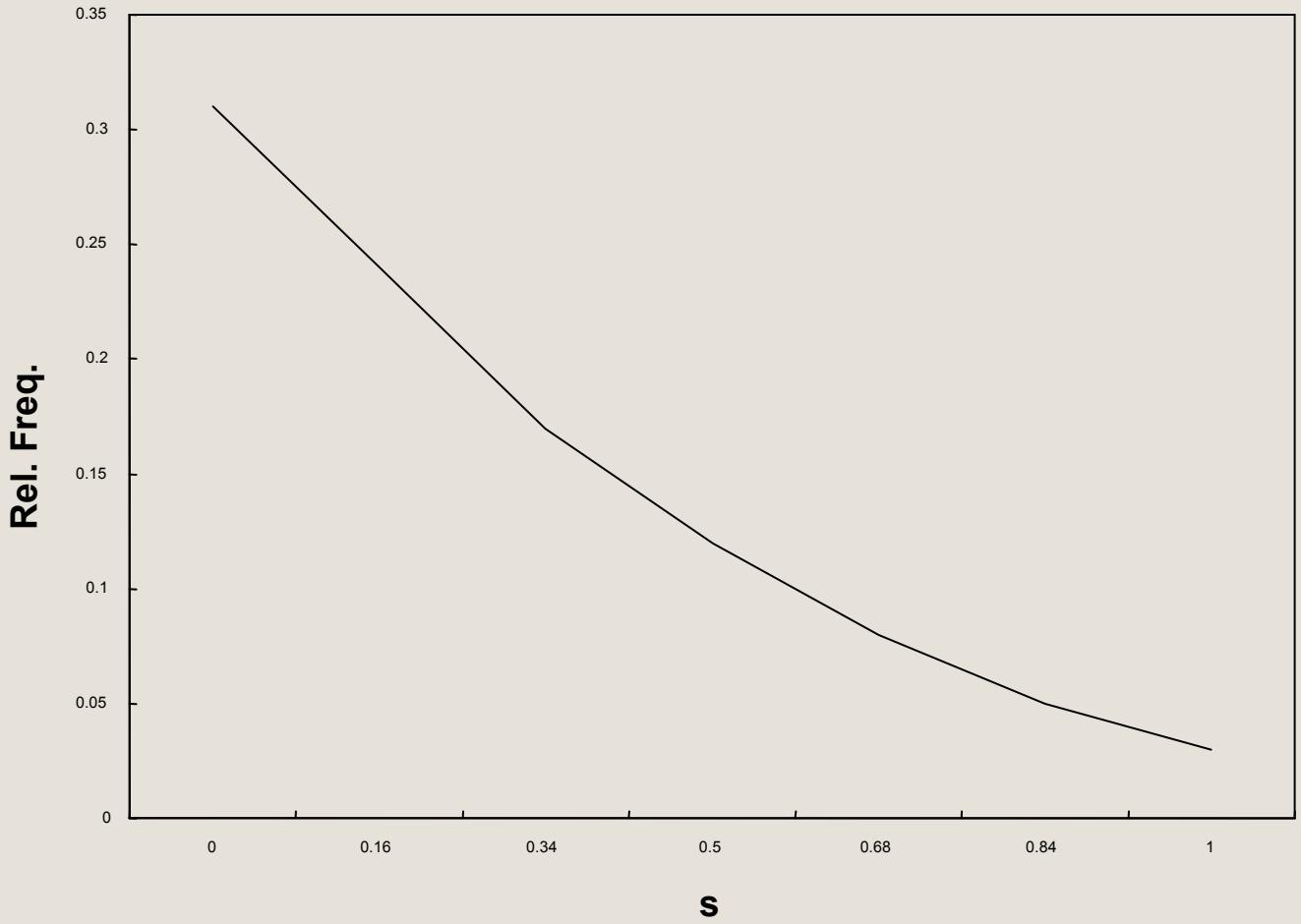
Fuzzy Signal Detection and Vigilance

Shifts in 'signal' probability

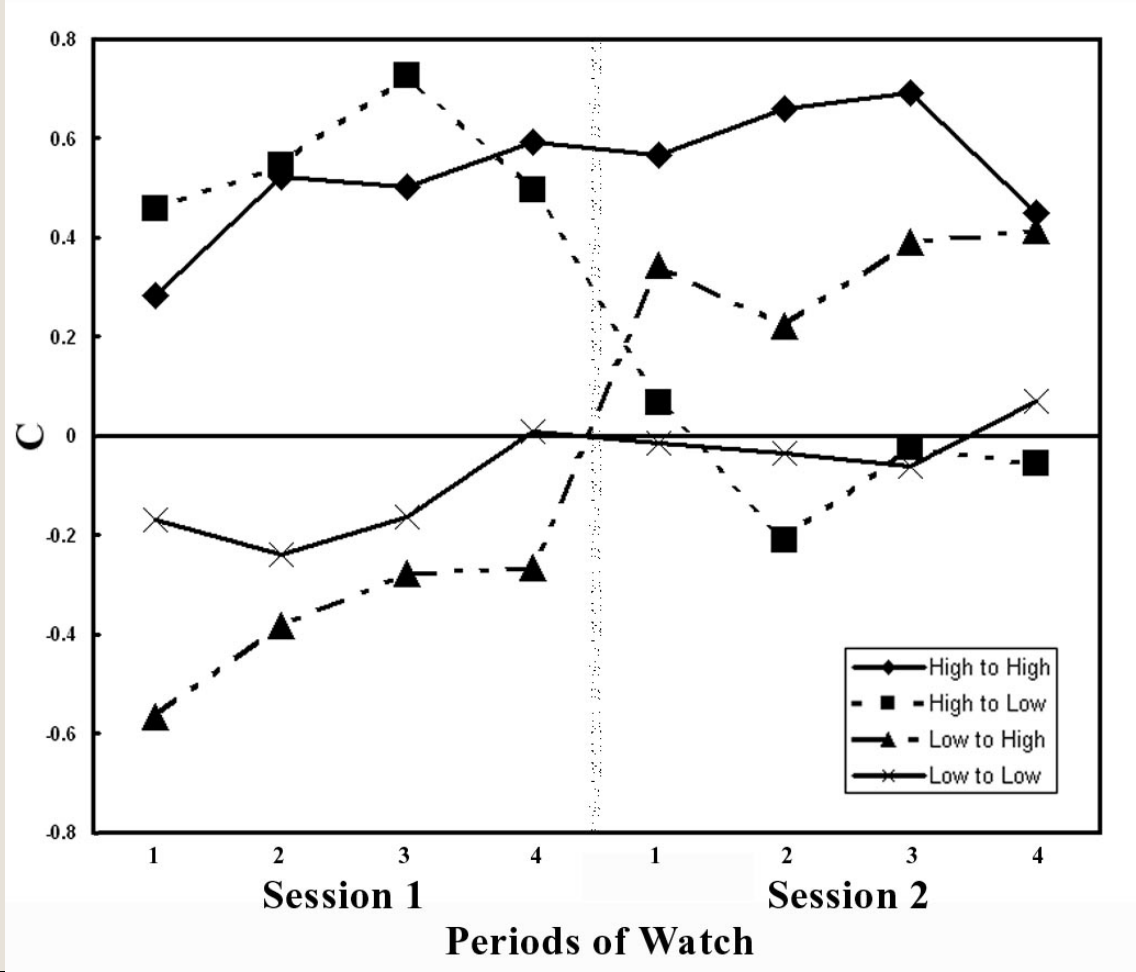
‘High’ Signal Probability



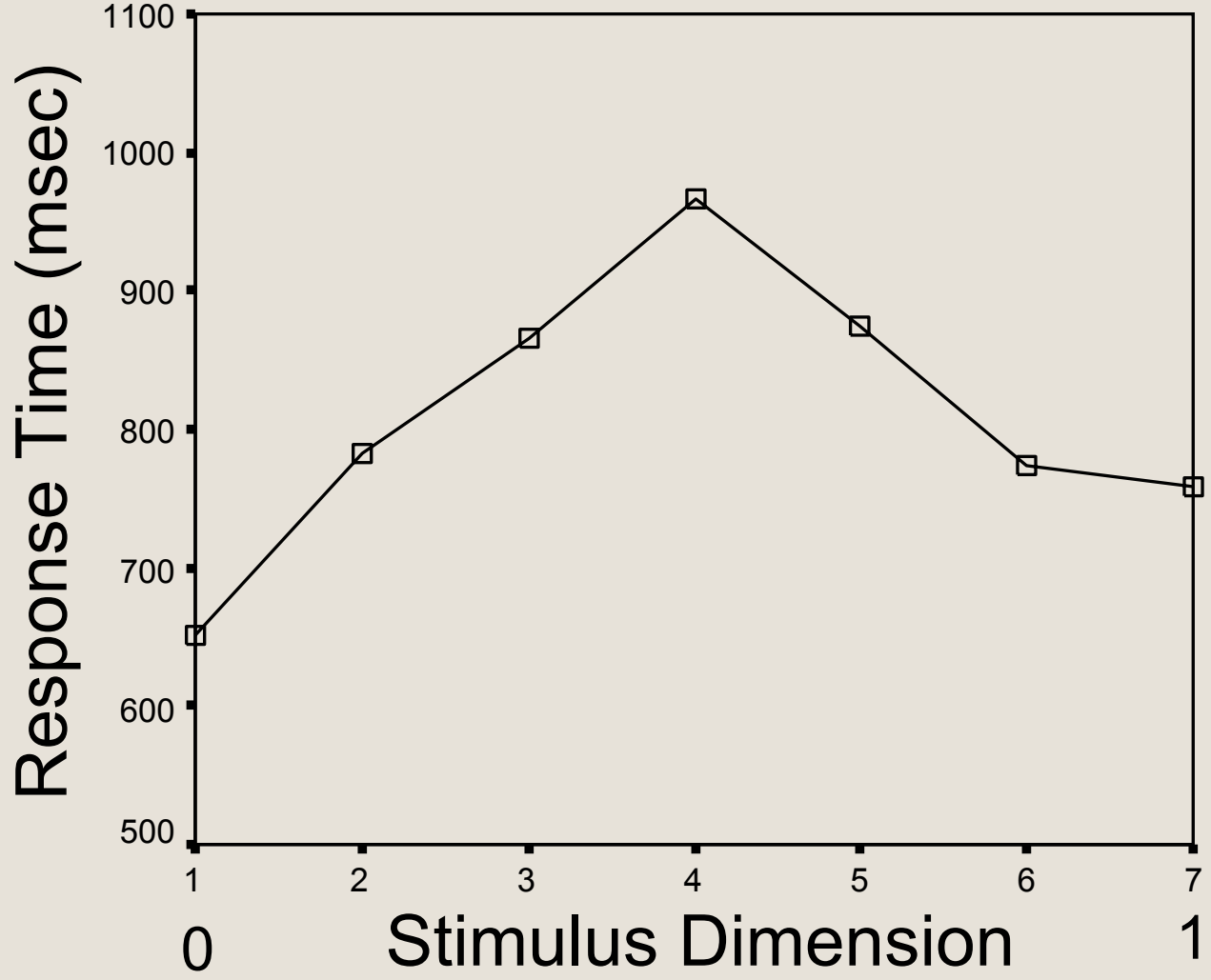
'Low' Signal Probability



Response Bias as a function of Periods of Watch



Response Time as a Function of Degree of Stimulus Criticality



ACKNOWLEDGEMENT

This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program , P.A. Hancock, principal investigator, administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the author and do not necessarily reflect official Army policy. The author wishes to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

The Role of Stress and Individual Differences in Warning Compliance

Helmick, J.A., Weaver, J.L., & Burke, K.A.
University of Central Florida

Introduction

Much of the research that has been conducted regarding warnings and compliance has focused on aspects of the warning itself (e.g., type of signal word, color, placement, etc.) and/or person variables that influence behavioral compliance to the warning (e.g., age, gender, hazard perception, etc.). However, some of this research has yielded equivocal results. Consequently, Weaver, Gerber, Hancock, and Ganey (in press) explored the role of sensation seeking and health orientation as individual difference variables, in addition to the effects of signal word presentation on behavioral compliance to a health warning. The results of the study provided support for these hypotheses in that those who were higher in sensation seeking required a stronger signal word to gain compliance. Also of interest, was the finding that individuals low in health orientation were less likely to comply in the "no warning" condition than those with higher health orientation. This data serves to support the general conception that individual difference factors such as health orientation and sensation seeking require consideration in evaluating the nature of warning compliance. In addition to these findings confirming that individual differences might play a significant role in determining the extent to which persons comply with warnings, they also indicate a potential explanation for past equivocal findings in warning research regarding the manipulation of signal word strength.

Consequently, the study reported here acts as a follow-up to Weaver et al. and seeks specifically to consider the role of arousal in relation to individual differences and compliance to warnings of varying strengths. In addition, the current study also investigates the above variables when participants are placed in stress (noise) vs. non-stress conditions. Finally, the study reported here also extends the work described in Weaver et al. (in press) by increasing the possible number of compliance behaviors and by utilizing a unique methodology (a wood-working task) in order to study the above variables. The following describes an ongoing study and provides preliminary results based on the first 30 of 120 participants to be observed.

It was hypothesized that warning compliance would be highest in the "Danger" condition and lowest in the control "Notice" condition. It was also hypothesized that low sensation seekers would comply when presented with a lower level of threat than would high sensation seekers. In addition it was hypothesized that high sensation seekers would report lower arousal upon arrival to the experiment and as assessed following task completion. Finally, it was expected that sub-scores on the NASA-TLX would vary across stressor condition and that compliance would be lower in the stressor condition.



Figure 1. Compliance gear and equipment that was available during the experiment: protective jacket, eyewash, rubber gloves, goggles, and face mask (Clockwise). The tool work station set-up (Center).

Method

Participants

Thirty (of the intended 120 participants) female University of Central Florida undergraduate students volunteered for this experiment in exchange for extra credit in an undergraduate psychology course. The participants were 30% freshmen, 33.3% sophomores, 30% juniors, and 6.7% seniors with a mean age of 19.67 (SD = 1.49). All participants were treated in accordance with the "Ethical Principles of Psychologist and Code of Conduct" (American Psychological Association, 1992).

Materials

Questionnaires:
- NASA-TLX (Hart & Staveland, 1988)
- Zuckerman's Sensation Seeking Scale-Form V (1994)
- Exit survey adapted from Braun's (1993) study on Color Product Warnings and Behavioral Compliance
- Otsubo's (Otsubo, 1998) behavioral study of warning labels and consumer products
- Demographic questionnaire
- Several questions about team performance

Equipment
- Small home surveillance camera
- Color video display terminal (VDT)
- Cassette of a table saw cutting wood at 85 decibels
- A Pioneer dual cassette player/recorder
- Two Bose speakers
- Sound pressure meter
- Dell Dimension 8200 desktop computer

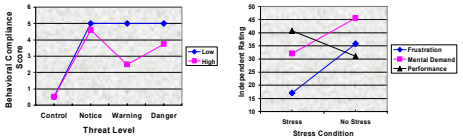


Figure 2. Warning compliance behavior for low and high sensation seekers as a function of threat level (Left).
Figure 3. Independent ratings of frustration, mental demand, and performance for Stress and No Stress conditions (Right).

*This research was facilitated by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ARO under grant DAAD19-01-1-0621.

Results

The dependent measures of interest were the behavioral compliance scores (the sum of the number of behaviors performed by each participant in each condition), Workload (as measured by the NASA-TLX), Sensation Seeking (as measured by Zuckerman's Sensation Seeking Scale-Form V), and Health Orientation (as measured by Bausell's Health Orientation Scale). For the analysis to be described, the between-subjects factors were the stress condition with two levels (noise, no noise) and the degree of threat of the warning with four levels (control, notice, warning, and danger).

Behavioral Compliance

The behavioral compliance score was obtained by summing the number of behaviors performed (donning: mask, gloves, goggles, and jacket) by each participant in each condition.

- A between-subjects 2 (low SSS vs. high SSS) X 4 (control, notice, warning, and danger) factorial ANOVA on behavioral compliance revealed a main effect of threat $F(3, 26) = 11.05$, $p < .0005$.
- Participants were less likely to comply with the warning at the Control condition ($M = .50$) than at the Notice ($M = 4.8$), Warning ($M = 3.75$), and Danger ($M = 4.38$) conditions.
- Although statistical significance was not reached, the main effect of SSS approached significance, $F(1, 26) = 3.29$, $p = .085$ with means of 3.88 and 2.84 for low and high sensation seekers, respectively, suggesting that with the exception of in the Control condition, high sensation seekers were less likely to comply with the warning than low sensations seekers.

NASA-TLX

A priori hypotheses were predicted with regard to the independent rating scales for Mental Demand, Performance, Effort, and Frustration. Several independent samples t-tests were conducted to identify if the noise stress affected the participant's subjective workload as measured by these rating scales and the weighted rating.

- *Although the weighted rating was not statistically significant, the independent scale rating for Frustration was significant, $t(28) = -2.04$, $p = .003$ (one-tailed).*
- Although not statistically significant, trends in the remaining independent scale ratings are consistent with predictionsA difference between means for the independent scale of Mental Demand (Stress: $M = 32.1$; No Stress: $M = 42.5$) as a function of stress.
- Participants in the Stress condition rated their performance as poor ($M = 40.8$), whereas participants in the No Stress condition rated their performance as high ($M = 31$).

Discussion

It is important to note that the results described here represent only a portion of the data that is currently being collected (i.e., 30 of 120 participants), thus, the results reported reflect only those observations. It is our belief that null results related to the manipulation check question, for example will likely be significant with stronger power.

Although there was not a significant difference in compliance across stress conditions, there was a significant finding for compliance as a function of threat, with the control condition being associated with the lowest levels of compliance. It is believed that this finding will also become more defined with the addition of more participants. That is, we would then expect to see differences in compliance between other levels of threat as well. Finally, although non-significant, the expected trend was present for sensation seeking and the interaction between sensation seeking and threat level. In conclusion, it is anticipated that the current study will replicate the findings of Weaver et al. (in press) and others in addition to making a contribution via the novel aspects of the investigation (e.g., the new methodology for the investigation of warnings).

The Application of a Commercially Available Software Program as a Testbed for Research on Stress and Performance.

Burke, K.A., Greenwood-Ericksen, A., & Weaver, J.L.

Due to the expense of obtaining access to purpose-built virtual environments, it has traditionally been difficult for some researchers to examine questions of interest involving infantry combat performance. As part of ongoing research into stress and human performance in infantry combat, a highly affordable soldier simulator was constructed using off-the-shelf parts and equipment. Simulators of this type are highly flexible and scalable in design, allowing researchers to fabricate them to the requirements of the study and/or in proportion to available resources. The current demonstration will allows participants to operate the simulator themselves, or observe the demonstrators operating it. The purpose of the demonstration is to acquaint human factors researchers with a system that demonstrates that solder simulator systems can be developed affordably and easily.

INTRODUCTION

There are many problems facing human factors researchers in examining the area of infantry combat. Due to the difficulty and expense involved in field studies, the use of simulation provides the most accessible option for researchers examining research problems in this area (Pleban et al, 2002, Driskell & Salas, 1991). While government agencies can afford to invest millions of dollars in soldier-simulator development, relatively few academic institutions can afford the necessary price tag. As a result, little research has been performed in academic settings using these devices. This problem has been exacerbated by the tendency of government agencies to develop or contract private companies to develop highly specific, simulator packages, rather than more flexible and general simulations that could be adapted for use by other researchers in a cost effective manner. Additionally, such systems are usually very technically complex and require high levels of technical competence to operate. The result is that many academic institutions are unable to afford the facilities or dedicated staff to conduct this type of context-contingent research.

However, it is possible to build more affordable, simpler substitutes for this type of simulator from easily available and relatively inexpensive “off the shelf” components. The development and examination of the design and capabilities of such a simulator is the purpose of the proposed demonstration.

The present work demonstrates a soldier simulator system assembled for under $6000 from such off-the-shelf components.

DESIGN

System Requirements

The purpose of the present project was to develop a medium-fidelity soldier simulator that would allow the researcher access to large numbers of dependent variables and performance metrics. Other requirements were that the simulator would preserve a high degree of face validity and adhere reasonably well to actual combat tactics and behavior. Additionally, the simulator needed to have good single and multiple simultaneous user capabilities, have an accessible scripting utility, and, most importantly, be affordable for small laboratories and universities.

The Software Package

The heart of any simulator is the software package that powers it. This role was to be filled by the use of a commercially available video game. In order to determine which of the wide variety of games available on the market would be the most appropriate choice, we surveyed and tested approximately a dozen alternatives. Ultimately, the

game “Tom Clancy’s Ghost Recon®” was chosen as the best choice for our software component. While the video game was critically acclaimed, receiving several industry and magazine awards for best PC game of the year, the choice was based largely on the factors which affected its performance as part of a soldier simulator. The simulator has a powerful and relatively accessible mission editor, allowing researchers a high degree of control over the environment and computer controlled actors. It has good face validity and weapons modeling, and enemy and friendly forces that behave in a reasonably realistic manner in the context of the simulation. The dynamics of the game force the user to consider the effects of their own movement, weapon choice, accurate range, and the use of cover and concealment by themselves and their enemies in order to effectively engage the enemy. The game is designed to reward good tactical behavior and punish unrealistic “run and gun” tactics common to first person shooter video games, in which the user races frantically about in the game environment, shooting up the landscape indiscriminately. Thus, game dynamics emphasize the use of proper tactical approaches to problems. Finally, the game is not only designed for single player use, but has well developed multiplayer capabilities for up to 36 networked human operators, who can be physically located as close as the next room or as far as halfway around the globe.

Hardware Components

The physical components of the simulator represent the bulk of the cost of the system. The prototype model used a digital video projector, a projector screen, and a PC equipped with a Pentium 4 with a 2.0 GHz processor, 256 megabytes of RAM, and an nVidia GE Force 4 graphics card with 64 megabytes of RAM. The PC system and the projector chosen largely determine the quality of the simulator. Less expensive models, made from less sophisticated parts, will produce a less graphically impressive environment, while higher quality hardware can boost resolutions up to near DVD quality resolution in 32 bit (True) color. With the best hardware currently available and the highest software settings, the simulator can project high-resolution imagery at almost life size. Sound is another factor in which great variation in quality is possible. For the purposes of maximizing immersion, the current system was designed with very advanced sound capabilities (an Extigy sound card by Audigy and Klipsch 5.1 surround sound speakers). The software component supports full 360 degree surround sound, so this relatively inexpensive feature not only greatly enhances immersion, but supplies task-relevant detail to the user. Additionally, investing in good quality sound provides good value for its expense, as it is quite important in developing and maintaining immersion (Stanney, 2002).

CAPABILITIES

The system’s capabilities are quite impressive for its low cost. The scripting and mission design tool is provided with the software and is a menu driven utility that is both powerful and easy to learn. This means that researchers themselves can quickly and easily learn to design their own scenarios without having to rely on a specialist programmer. Through the editor, users can control almost every aspect of the combat environment, from lighting to weather, and can script them to change dynamically throughout the course of the mission. Explosions or airstrikes can be simulated at certain locations at certain times or in response to player actions. Additionally, the user can control the positions, armament, appearance, disposition, and behavior of enemy and friendly units in as little or as much detail as necessary. This means that the experimenter can control the number of friendly and enemy units, their equipment and armament, their training and quality, and their general behavior in combat. Enemy units can be scripted to panic under fire, use cover or concealment, set ambushes, operate vehicles, and work together effectively (or ineffectively if so scripted) as a team.

Monitoring Events and Data Collection

One of the major strengths of the system is the freedom it allows the user to monitor events in the simulation space. The experimenter can actually enter the mission environment visibly or invisibly to observe events in real time by means of the

multiplayer functionality. A replay feature logs all control inputs made by the user and tracks the position and behavior of every other character and element in the simulation. This feature allows the experimenter to replay the mission from multiple angles and multiple viewpoints after the participant has finished the task. Replays can be saved for future review. Additionally, the simulation itself keeps track of several important statistics on a running basis, including number of shots fired, number of hits made, number of kills, shooting accuracy, number of squadmates hit or killed, number of mission objectives completed, in progress, or failed, and the amount of time elapsed since the simulation began. This information is displayed following the completion of every mission in a debriefing report.

Simulator Limitations

The simulator does have some limitations, however. The most glaring of these is the inability to interface with a surrogate weapons system or any elaborate movement control. Since it is designed exclusively as a video game, there is no way to avoid using some kind of PC-based input device. Thus, the simulation must always be based around a keyboard, joystick, mouse, or similar input system. Thus there is no provision available for the simulator to interface with a surrogate weapons system such as many advanced dismounted infantry simulators use (SAST, DISALT, others). Additionally, because of the wide array of behaviors available to the user in the simulator, the interface is by necessity fairly complex, and is somewhat non-intuitive in design due to the limitations of the input devices. We have experimented with alternate input devices and layouts, but the problem still remains difficult to solve without at least 10-15 minutes of subject training. Another shortcoming of the simulator is the limitation of the number of friendly soldiers under the user's control allowed in a single player mission. The game sets an arbitrary limit of six squadmates under the user's direct control in single player mode, divided into up to three fire teams. Additionally, the largest fire team possible is three, so in order to use all six possible squadmates in single player mode, it is necessary to use two independent fire teams. This might pose difficulties in coordination for less experienced users.

In general, however, we think that the present system has great potential for researchers interested in pursuing research questions related to behavioral aspects of infantry combat without the expenditure of large amounts of money and time. Potential applications for this simulator system in research include navigation, target detection or identification, tests of transfer of training, use of tactical strategies, and tests of short term memory and stress effects.

ACKNOWLEDGEMENTS

This research was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

REFERENCES

Driskell, J.E., & Salas, E. (1991). Overcoming the effects of stress on military performance: Human factors, training, and selection strategies. In Gal, R. & Mangelsdorff, A.D. [Eds] *Handbook of Military Psychology*. New York: John Wiley and Sons.

Pleban, R.J., Matthews, M.D., Salter, M., Eakin, D.E. (2002). Training and assessing complex decision making in a virtual environment. *Perceptual and Motor Skills, 94*, 871-882.

Stanney, K. [Ed]. (2002). *Handbook of virtual environments: Design, implementation, and applications.* Lawrence Erlbaum Associates, Publishers, Mahwah, NJ.

ERGONOMICS IN DESIGN

The Magazine
of Human
Factors
Applications

Operator Stress & Display Design

BY PETER A. HANCOCK & JAMES L. SZALMA

Reprinted from *Ergonomics in Design,* Volume 11, Number 2, Spring 2003; pages 13–18. Copyright 2003 by the Human Factors and Ergonomics Society, P.O. Box 1369, Santa Monica, CA 90406-1369; 310/394-1811, FAX 310/394-2410, info@hfes.org, http://hfes.org. All rights reserved.

Operator Stress and Display Design

These design guidelines, based on knowledge of how operators perform under time and task pressure, can lead to better decision making in emergencies.

BY PETER A. HANCOCK & JAMES L. SZALMA

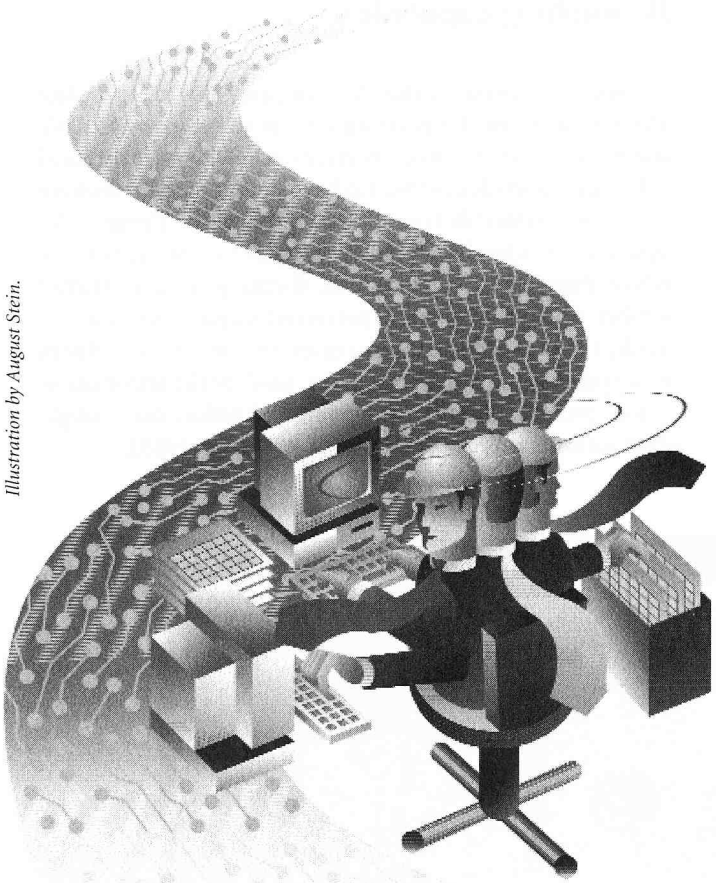


Illustration by August Stein.

IT IS NEAR THE END OF A LONG SHIFT, some hours past midnight, and the combined effects of caffeine, nicotine, and sugared snacks are no longer sufficient to sustain the alertness of tired and weary operators. Circadian rhythms are at their lowest ebb while oppressive, routine operations have so long persisted that chronic boredom reigns supreme. Suddenly a warning sounds, then another, then another – the systems displays are producing a cascade of flashing lights and tone alarms.

Operators rush to workstations, where information begins to overflow like a virtual Niagara of bytes. The apprehension in the air is tangible. It is evident to everyone that a coherent response must be made soon if the situation is to be recovered. As a result, time pressure is now added to anxiety and information overload and rapidly shifts the situation from boredom to terror (Hancock, 1997a). The answers to the problem are somewhere in the system, but a massive noise-to-signal problem is building by the minute. Unfortunately, the resolution is spread across a tapestry of displays. Like pieces of a dynamic jigsaw puzzle, the solution resides in separate parts distributed across many minds, but team communication and team cohesion are rapidly failing, overwhelmed by the demands of information overload and the destructive effects of situational stress. If someone doesn't do something soon . . .

The foregoing account is fictitious, but it might well represent any number of real-world emergency experiences in circumstances as diverse as a nuclear power station control room, a military command and control facility, or the nerve center of a large urban emergency response team. Operators across many application domains are required to perform under situational stresses. However, the problem of information overload is not restricted to emergency circumstances alone but is the common experience of most computer users. Time pressure is also a ubiquitous imposition and not restricted to critical situations.

Observations such as these confirm the growing consensus that the dynamic context of performance is a crucial arbiter of design (Flach, Hancock, Caird, & Vicente, 1995). The premise that many jobs have to be performed in the presence of high task workload and situational stress is our focus in this article. A number of observations about humans performing under workload and stress permit the derivation of some general principles or guidelines for display design for these conditions. Here we enumerate some of these guidelines.

The Stressed Operator

What is the best approach to attack the problem of designing for stressed operators? If consistent and informative principles already existed, one might be able to answer this question at an overall systems design level. However, the present state of affairs debars this high level of analysis because the basic knowledge is not yet well articulated. In this case, one needs first to consider how individual operators react to high-stress and high-workload conditions. Fortunately, there is some useful information regarding these forms of behavioral response (see Hancock & Desmond, 2001). It is known, for example, that extremes of stress affect sensory and perceptual capacities, which results first in a diminution and then a failure of information assimilation. Such degradation denies the subsequent opportunity to engage in the appropriate decision-making and response execution.

In essence, when anticipating high-stress situations, designers need to be aware that they are dealing with selective and diminishing capacities.

Figure 1 represents the changing capacity to assimilate information from the environment as stress increases. We base these observations on an extensive review of theoretical and empirical evidence that we have documented elsewhere (Hancock, Szalma, & Weaver, 2002). Primarily through the depletion of attentional capacity, operators systematically reduce their information intake, focusing on a restricted number of cues of greatest perceived importance (Easterbrook, 1959). As the stress increases, the individual reduces the number of information sources, until, at the termination of this process, there is an apparent fixation on a single informational source (see Hancock & Dirkin, 1983).

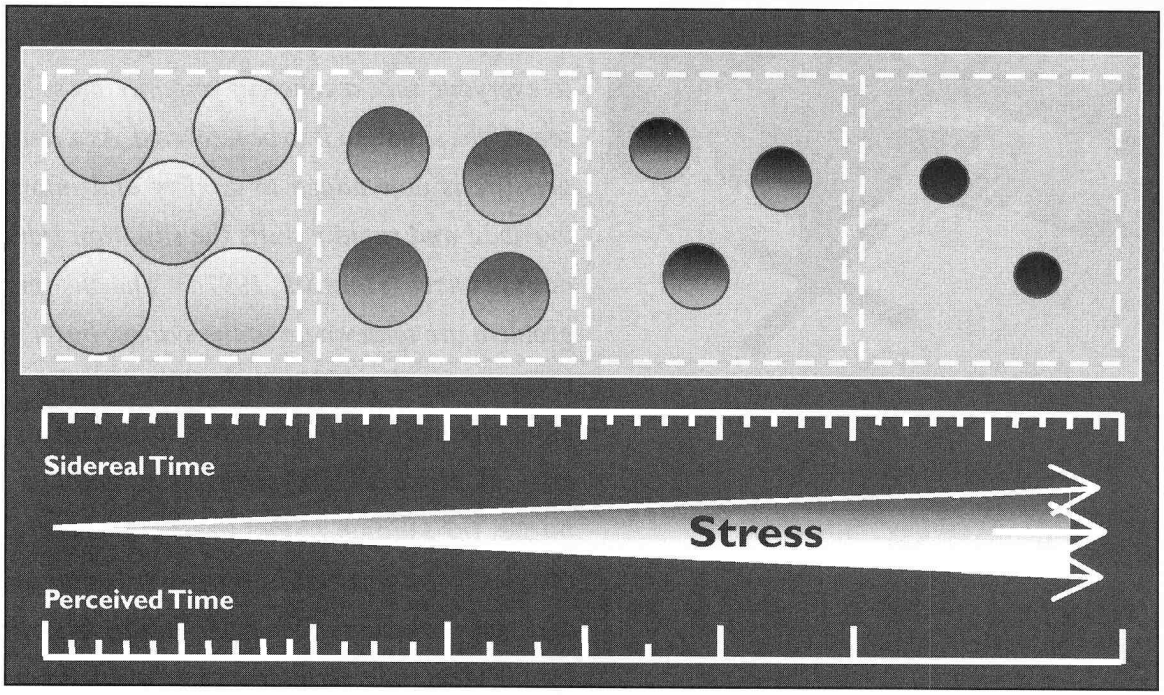


Figure 1: Systematic distortion of perceptual space-time with increase in stress. The upper panel shows a reduction in scan range together with greater focus on each fixation. The lower panel (bracketing the stress level) shows that temporal perception is affected in a similar manner to the spatial distortion.

A little-recognized correlate of this activity is the reduction of time sampling. In addition to variation in spatial sampling, individuals experience significant distortions of time as their restricted attention is diverted away from temporal cues. This distortion is especially important when timing is a critical facet of the task at hand (see also Hancock & Weaver, 2003). This process is a perceptual form of narrowing, which we suggest is replicated at a macro level in strategic task shedding. It has been a source of contention whether the narrowing phenomenon is largely perceptual or sensory in nature. Evidence suggests that operators are able to narrow to salient cues that appear in any part of the visual field (Dirkin & Hancock, 1985; Hancock & Dirkin, 1983; Hockey, 1970, 1983). Narrowing, having been replicated in other sensory capacities (Bacon, 1974), can be taken as an attentional strategy engaged by higher centers of the brain. We have recently proposed that the distortions of both space and time result from a common attentional mechanism (see Hancock et al., 2002).

Much of the narrowing effect is automatic and under only very limited voluntary control, although through adaptive behavior, the stressed operator can effectively accomplish a number of actions before dire effects prevail. Typically, one expects to see conscious load-shedding prior to this narrowing stage of stress response. Also not completely confined to conscious attention, shedding is the action of rejecting irrelevant or, at that moment, unwanted tasks or sources of information. In teams, shedding may take the form of task-sharing or task redistribution. In a single human-machine dyad, shedding might be embedded in the original task allocation design policy. In adaptive human-machine systems, task reallocation is designed to occur as either the human or the machine approaches close to his/its respective capacity limits (see Hancock, Chignell, & Lowenthal, 1985; Hancock & Chignell, 1988). Each of these forms of shedding requires some form of communication, often manifested in the observation that a particular task component is no longer being performed up to standard.

Task demand level can be changed by design, and so stress itself may actually come under the designer's control.

In essence, when anticipating high-stress situations, designers need to be aware that they are dealing with selective and diminishing capacities. Moreover, these capabilities diminish in a systematic fashion as more demanding tasks apparently drain resources at a faster rate than less demanding tasks. Similarly, tasks sharing common resource capabilities, such as two coincident tracking tasks, are more vulnerable to the effects of stress than are tasks that do not have significant overlap, such as a dichotic listening task combined with a tracking task (see Szalma & Hancock, 2003; Wickens, 1984). Expecting stressed operators to seek and distinguish novel sources of information is a fallacy that should be avoided by designers, as we indicate in our design recommendations at the end of this article.

Task Stress and Display Design

For many decades, stress was conceived overwhelmingly as a property of the surrounding environment. The stress literature is redolent with experiments investigating the effects of temperature, noise, vibration, and other factors on task performance (Broadbent, 1971; Poulton, 1970). Along with characteristics of the environment, operator states such as fatigue, lack of sleep, and drug levels also affect performance capacity in a similar, often deleterious manner (Hockey, 1983).

What has been emphasized in more recent theories of stress is that the task itself is the proximal form of stress (Hancock & Warm, 1989). That is, among the various sources of stress, the demands of the task at hand often place the highest level of stress on an individual. As with stress in general, these effects are most obvious at the extremes where the demand generates significant overload or underload on the resource-limited operator. Task demand level can be changed by design, and so stress itself may actually come under the designer's control.

One of the significant issues faced by the designer looking to support decision makers who act under stress concerns the problem of integrating new information. As noted earlier, stress acts to concentrate the individual's attention on a restricted number of sources with perceived high reliance. Further, stress exacerbates the dependence of the decision maker on scenario fulfillment derived from his or her previous experience and current mental model of the situation. Such influences mean that stressed individuals are very likely to discount new information or information that runs contrary to their expectations (Hancock & Mortimer, 2003). The challenge to the designer is how to bring such critical information to the immediate attention of the user.

As increasing levels of stress are encountered, the transmission of complex alphanumeric information is likely to become a further source of distraction, rendering displays uninformative and perhaps deleterious in their performance effects. In extremis, when the person is faced with a time-critical, life-or-death decision, it is important to frame information to address the lowest level of processing capacities. Such displays are likely to be graphic representations of material restricted to an extremely low baud rate with simple messages of affirmative command (e.g., a green arrow pointed toward safety, a red X to indicate danger). When at the very extremes of their tolerance, operators are likely to ignore complicated information regardless of the modality of presentation or even its task salience.

One can, however, take one step back from this edge of incipient failure to look at the process of information integration to the stressed but not terminally challenged user.

This is important because our recent experimental findings have shown that professional operators can be pushed almost to the edge of resilience without a significant decrease of operational capacity (Harris & Hancock, 2003). Our data on stress effects show the presence of this critical failure level in which individuals continuously employ their adaptive and resilient capacities up to the edge of resistance, after which rapid performance failure occurs (see the elbow of the extended–U description of Hancock & Warm, 1989).

Expecting stressed operators to seek and distinguish novel sources of information is a fallacy that should be avoided by designers.

The task of the designer is to support the individual in this phase before the point of incipient failure. Our recommendation for this support, derived from a number of experiments, is that designers should use integrated displays in which different facets of system information are brought into a coherent whole so that the operator's need to engage in information integration is dramatically reduced. Such displays can be beneficial to performance (see Bennett & Flach 1992; Sanderson, Flach, Buttigieg, & Casey, 1989) and, under certain conditions, can reduce mental workload (Szalma, 2002).

Specifically, in a monitoring task requiring information integration, Szalma (2002) reported that object displays were more effective in reducing the perceived workload than were bar graph displays configured to produce an emergent feature representing global system state. It is our contention that these design preferences adhere to the tenets of ecological interface design (EID), which we see as appropriate for the support of stressed and overloaded operators (and see Vicente, 1999). Experimental work is still needed to fully establish exactly how workload and stress states are moderated by use of these displays. Szalma's (2002) recent results suggest that these effects are task and context dependent.

If distortions of spatial and temporal perception resulting from attentional narrowing represent a central characteristic of the stressed operator, displays should be designed to simplify the perception of these dimensions. Integrated displays accomplish this for spatial distortions by centralizing information in a single figure in an easily perceived format. To compensate for temporal distortion, these displays, like trend displays, should incorporate changes in system state as a function of time as a part of the collective integration process. Capture of the temporal dimension into an integrated display can permit the extraction of information regarding system state over specified time intervals (see Figure 2).

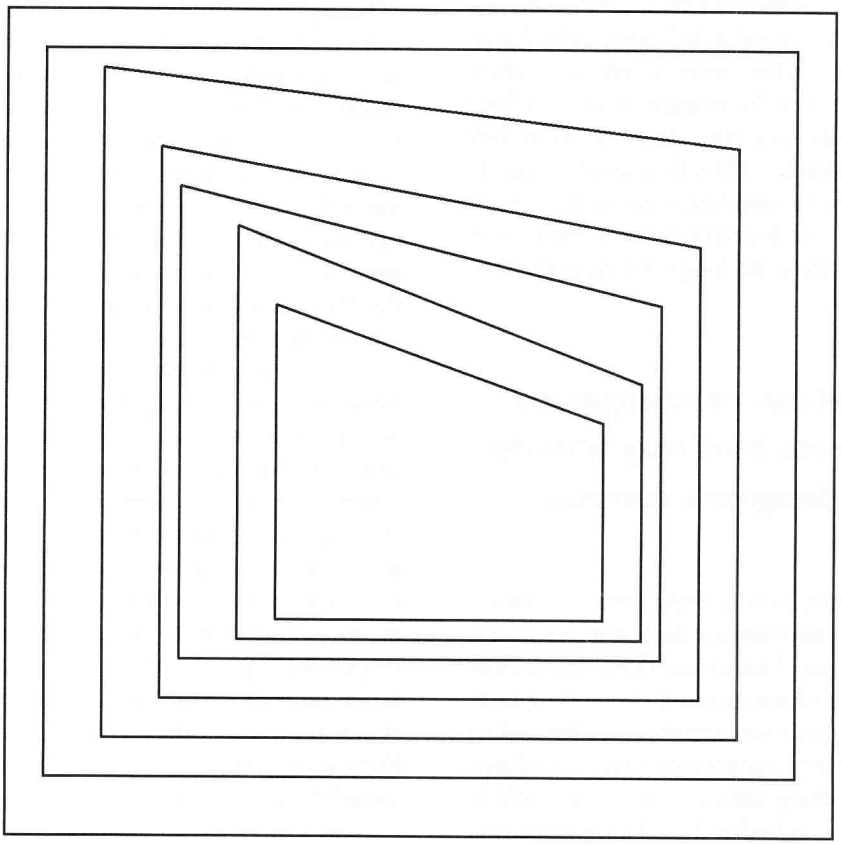


Figure 2. An integrated display organized in a time tunnel format. The successive changes in the size of the polygon represent the state of the system at different points in time. Variations, including confidence intervals, can be developed from this basic illustration.

Fortunately, data support the design use of this approach. For example, Hansen (1995) investigated the efficacy of incorporating temporal information into integrated process control displays. He reported that despite the contextual dependence of performance with configural displays, incorporating a temporal component into an integrated display can improve the ability of operators to perceive trends in the presence of display disturbances.

Designers should use integrated displays in which different facets of system information are brought into a coherent whole.

Given the finding that this display format aided performance under degraded viewing conditions, it is reasonable to infer that it will also support performance under high-stress conditions. Hence, it is possible to develop integrated displays with both spatial and temporal elements in a simplified arrangement that will mitigate the combined effects of distortions of perceptual space-time. However, as in the case of configural and integrated displays generally, it is crucial that the temporal changes represented in the display possess good semantic mapping to the temporal component of system states if such display formats are to be effective (Bennett & Flach, 1992).

Design Guidelines

Although operators in general do not face the problem of decisions under the extreme stress of crises such as the Three Mile Island and Blackhawk down incidents, they often experience periods of particularly high demand combined with limited time for response. In contrast, many workers suffer under the "hurry up and wait" situation in which prolonged periods of underload are followed by moments of mayhem (Hancock, 1997a). In addition, some military and emergency response personnel have to operate under conditions of ultimate threat.

In all these circumstances, effective display design can support performance, but poor displays can hamper and distract an already taxed individual. As a consequence, we have distilled the following guidelines to support display design for operations under stress.

- If at all possible, designers should minimize information dispersal over multiple sources for work under stress. This is because high-stress and high-workload conditions inhibit operator search for novel sources of information. This applies especially to situations in which task resolution requires information integration.
- New information should be linked to data currently being processed, given that operators under stress and high workload tend to discount new information.
- Whenever possible, designers should employ integrated information formats to present an overall "picture" on a single display. This also facilitates the integration of new or contradictory information. When appropriate, display elements should also capture temporal change to mitigate potential deleterious effects of temporal distortion.
- When integrated displays (e.g., object displays) are not practical, use separable configural displays (e.g., bar graph displays) whose elements (both spatial and temporal) form emergent features that are mapped to system dynamics (i.e., displays with good semantic mapping).
- The principles of ecological interface design facilitate implementation of the aforementioned principles. Thus, designers should explore further applications of EID to highly demanding operational conditions.
- When operators are faced with the most extreme conditions, designers should avoid presenting displays that require data transformation. The use of simple graphic displays to convey low-level, direct instructions is recommended because stress depletes resource-based processing of higher-level functions (e.g., problem solving).
- When possible, designers should first prevent direct structural interference to highly stressed and loaded operators. For example, when auditory information is masked by noise and visual displays are obscured by glare, failure to address these simple structural forms of interference obviates any subsequent improvement regardless of the level of sophistication of the design.

Conclusion

Designing for stressful and high-workload situations is becoming increasingly important with the continuing growth of the human as a system monitor and overseer. Specifying crucial information during violent swings between extremes of underload and overload is a vital concern because temporal distortion and time criticality characterize these life-altering events (Huey & Wickens, 1993).

Designers also have to be concerned with person-centered issues. Designing displays for the presentation of information may seem to be a task-driven endeavor, but in reality it is concerned with individual human intention and purpose (Hancock, 1997b; Oborne, Branton, Leal, Shipley, & Stewart, 1993). Indeed, the experience of stress and overload is directly contingent on the understanding of the exposed individual and the way he or she appraises the work environment in the first place (see also Lazarus & Folkman, 1984).

In giving the current recommendations, we acknowledge that much remains to be accomplished to refine the present guidelines and to integrate them with the few extant theories of interface structure and function. However, the move toward context-specific design is an important development, and we see this work as one step in that direction.

References

Bacon, S. J. (1974). Arousal and the range of cue utilization. *Journal of Experimental Psychology, 102*, 81–87.

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34*, 513–533.

Broadbent, D. E. (1971). *Decision and stress*. London: Academic.

Dirkin, G. R., & Hancock, P. A. (1985). An attentional view of narrowing: The effect of noise and signal bias on discrimination in the peripheral visual field. In I. D. Brown, R. Goldsmith, K. Coombes, & M. A. Sinclair (Eds.), *Ergonomics International 85: Proceedings of the Ninth Congress of the International Ergonomics Association* (pp. 751–753). London: Taylor & Francis.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*, 183–201.

Flach, J., Hancock, P. A., Caird, J. K., & Vicente, K. (Eds.). (1995). *Global perspectives on the ecology of human-machine systems*. Mahwah; NJ. Erlbaum.

Hancock, P. A. (1997a). *Hours of boredom, moments of terror – or months of monotony, milliseconds of mayhem*. Paper presented at the Ninth International Symposium on Aviation Psychology, April, Columbus, OH.

Hancock, P. A. (1997b). *Essays on the future of human-machine systems*. Eden Prairie, MN: Banta.

Hancock, P. A., & Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man and Cybernetics, 18*, 647–658.

Hancock, P. A., Chignell, M. H., & Lowenthal, A. (1985). KBAM: A prototype knowledge-based adaptive man-machine system. In I. D. Brown, R. Goldsmith, K. Coombes, & M. A. Sinclair (Eds.), *Ergonomics International 85: Proceedings of the Ninth Congress of the International Ergonomics Association* (pp.103–105). London: Taylor & Francis.

Hancock, P. A., & Desmond, P. A. (Eds.). (2001). *Stress, workload and fatigue*. Mahwah; NJ: Erlbaum.

Hancock, P. A., & Dirkin, G. R. (1983). Stressor induced attentional narrowing: Implications for design and operation of person-machine systems. *In Proceedings of the Human Factors Association of Canada* (vol. 16, pp. 19–21). Mississauga, Ontario: Human Factors Association of Canada.

Hancock, P. A., & Mortimer, D. C (2003). *Decision errors under stress*. Manuscript under review.

Hancock, P. A., Szalma, J. L., & Weaver, J. L. (2002). *The distortion of perceptual space-time under stress*. Paper presented at the 23rd Army Science Conference, Orlando, FL, December. (Complete White Paper version available at www.mit.ucf.edu).

Hancock, P. A., & Warm, J. S., (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519–537.

Hancock, P. A., & Weaver, J. L. (2003). Temporal distortions under extreme stress. *Theoretical Issues in Ergonomic Science*, in press.

Hansen, J. P. (1995). An experimental investigation of configural, digital, and temporal information on process displays. *Human Factors, 37*, 539–552.

Harris, W. C., & Hancock, P. A. (2003). *Field assessment of cognitive performance*. Manuscript submitted for publication.

Hockey, R. (1970). Signal probability and spatial location as possible bases for increased selectivity in noise. *Quarterly Journal of Experimental Psychology, 22*, 37–42.

Hockey, R. (Ed.). (1983). *Stress and fatigue in human performance*. Chichester, England: Wiley.

Huey, B. M., & Wickens, C. D. (Eds.). (1993). *Workload transition: Implications for individual and team performance*. Washington, DC: National Academy Press.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer-Verlag.

Oborne, D. J., Branton, R., Leal, F., Shipley, P., & Stewart, T. (Eds.). (1993). *Person-centred ergonomics: A Brantonian view of human factors*. London: Taylor & Francis.

Poulton, E. C. (1970). *Environment and human efficiency*. Springfield, IL: Charles C Thomas.

Sanderson, P. M., Flach, J. M., Buttigieg, M. A., & Casey, E. J. (1989). Object displays do not always support better integrated task performance. *Human Factors, 31*, 183–198.

Szalma, J. L. (2002). Workload and stress of configural displays in vigilance tasks. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 1536–1540). Santa Monica, CA: Human Factors and Ergonomics Society.

Szalma, J. L., & Hancock, P. A. (2003). *On mental resources utilization under stress*. Unpublished White Paper, University of Central Florida, Orlando, FL. (available at: www.mit.ucf.edu).

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). Orlando: Academic.

Peter Hancock is Provost Distinguished Research Professor in the Department of Psychology, the Institute for Simulation and Training, and at the Department of Civil and Environmental Engineering at the University of Central Florida, phancock@pegasus.cc.ucf.edu. He is the principal investigator on the Multidisciplinary University Research Initiative, Operator Performance Under Stress (MURI-OPUS), which is a five-year project examining the effects of stress, workload, and performance on soldier performance. His current experimental work concerns the evaluation of behavioral response to high-stress conditions. James L. Szalma is a postdoctoral research associate in the Psychology Department and the Institute for Simulation and Training at the University of Central Florida. His current research interests include workload, stress, and motivational issues in human performance, individual differences in performance, and in the way humans experience their interaction with technology.

This research was supported by a grant from the Army Research Organization (#DAAD19-01-1-0621), P. A. Hancock, principal investigator. The work was facilitated by a grant from the U.S. Army to the Advanced Decision Architecture Consortium (#DAAD19-01-2-0009), P. A. Hancock, principal investigator. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the U.S. government. The authors thank Elmar Schmeisser, Sherry Tove, Mike Drillings, Paul Gade, Rene De Pontbriand, and Laurel Allender for providing administrative and technical direction for the first grant.

EFFECTS OF AUTOMATION RELIABILITY ON HUMAN MONITORING PERFORMANCE

Brian Oakley, Mustapha Mouloua, & Peter Hancock
University of Central Florida
Orlando, FL

Human decisions to rely on automation of a simulated flight task were monitored in the present study. Participants performed the Multi Attribute Task (MAT) battery previously developed by Comstock and Arneguard (1990). Twenty-eight participants were tested for 7 different reliability levels. Decision accuracy, false alarm rates, and reaction time to simulated automation failures were collected for each participant over nine 10-minute sessions. The results indicated that the rate of automation failures detected varied inversely with automation reliability level. As automation level increased, detection rate decreased thereby indicating automation-induced monitoring inefficiency. However, the mean reaction time of correct detection rate decreased as automation reliability level increased. No significant effect of reliability was obtained for the false alarm rate data. The implications for training and systems design are also outlined.

INTRODUCTION

The implementation of advanced automation technology in aviation systems has resulted in several benefits (e.g., all-weather flying, increased safety, fuel efficiency, decrease in physical workload, etc.,). However, there are several automation costs that have also been attributed to increased automation use. These behavioral problems include automation-induced monitoring inefficiency, loss of situation awareness, degradation of manual flight skills, inability to revert to manual control in case of automation failures, and increased mental workload. According to the Aviation Safety Reporting System (ASRS), there are many instances in which monitoring failures were linked to excessive trust in or over reliance on automation such as the autopilot or FMS (Mosier, Skitka, and Korte, 1994).

In general, human monitoring of automated systems for possible malfunctions or failures has resulted in increased demands on human operators (Chambers and Nagel, 1985; Parasuraman, 1987; Wiener, 1988). An extensive body of literature has well documented several aspects of these behavioral problems across different domains of applications (Mouloua and Parasuraman, 1994; Parasuraman and Mouloua, 1996; Mouloua and Koonce, 1997, Scerbo and Mouloua, 1999).

Research on vigilance has also shown that human detection of low-probability events is poor during sustained watch keeping operations (Davies and Parasuraman, 1982). Monitoring automated subsystems for automation failures has also been poor when automation is static (Parasuraman, Molloy, and Singh, 1993; Singh, Molloy, and Parasuraman, 1993). These results have been attributed to the performance costs associated with static automation. In contrast, adaptive automated tasks in which the allocation of function between the machine and operator is not static have been touted to be superior (Rouse, 1988, Parasuraman, Mouloua, and Molloy, 1996). Because most of these studies have used different reliability levels, it is not clear whether some of these effects can also be attributed to automation reliability.

Operator's trust in or over-reliance on automation is an important factor in determining when a system can be trusted to function appropriately and when a system is unreliable and trust is unwarranted (Muir, 1988). Subsequently, May, Molloy, and Parasuraman (1993) examined the effects of automation reliability on performance of an automated monitoring system in a pair of

experiments. While the first experiment reported that the detection rate of automation failures decreased as the level of automation reliability increased; experiment 2 failed to report any significant reliability effect with only two reliability groups. This could be because the range between the two groups was not wide enough and could have caused high individual differences (Singh, Molloy, & Parasuraman, 1993); thereby resulting in substantial error variance in experiment 2.

The present study was designed to further investigate factors affecting the establishment of automation trust in the context of monitoring performance. Automation reliability was varied over 7 different levels ranging from 5% to 95%. It was hypothesized that increased levels of automation reliability would result in degraded performance of the system-monitoring task.

METHOD

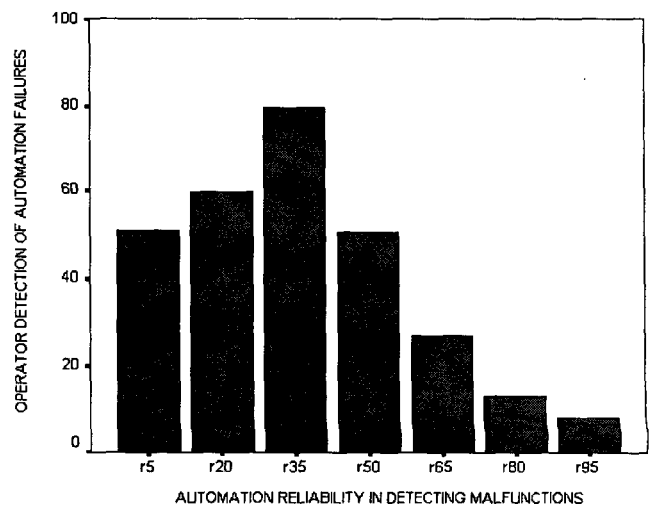
Twenty-eight volunteers participated in this study. A 7 (reliability) by 3 (session) mixed factorial design was used. The reliability of the automation system in the monitoring task was manipulated as a between-subjects *(5%, 20%, 35%, 50%, 65%, 80%, and 95% reliability)* factor. The within-subjects factor in the factorial design was the three 30-minute sessions. Automation reliability was defined as the percentage of the 16 system malfunctions correctly detected by the automation routine in each of the 10-minute blocks. A modified version of the Multi-Attribute Task Battery (MAT) developed by Comstock and Arnegard (1990) was used for each session. The revised MAT is comprised of three tasks: tracking, system monitoring, and resource management.

The system-monitoring task was automated to detect any offsets in which pointers on one of four dials went out of range. Upon detection, the malfunction was reset automatically. Periodically the automation system would fail to detect the malfunction. Upon this occurrence, the subject was instructed to manually correct the offset by pressing the corresponding function key. If 10 seconds elapsed without detection, the offset was automatically corrected. A first-order, two-dimensional compensatory tracking task was used as the second task. The task of the subject involved keeping an aircraft icon in the center of the tracking display. The fuel management task was designed to simulate the tasks performed in managing the fuel system of an aircraft. The subjects were instructed to compensate for fuel depletion by pumping fuel from the supply tanks to the main tanks. Root mean square (RMS) error was computed for both the tracking and fuel management tasks.

Subjects were instructed to perform all three tasks manually in a 10-minute training period and to give equal attention to each of the tasks. The subjects then began three 30-minute sessions, which were comprised of three 10-minute blocks. The subjects were informed that the system-monitoring task had been automated and their primary tasks were tracking and fuel management. In addition, they were also advised that the automation routine may not be 100% reliable and therefore they needed to monitor it for automation failures.

RESULTS AND DISCUSSION

Figure 1 shows the mean percentage of the automation failures detected for each of the seven groups. As the level of reliability increased, the detection rate decreased. An analysis of variance (ANOVA) of the detection rate showed a significant effect of automation reliability, $F(6,21)=5.88$; $p<.005$



An initial analysis of the data revealed no difference between the groups due to a resource trade-off effect between the three tasks of the MAT. In other words, a high percentage of detection of the automation failures could be obtained at the expense of the performance on the other two tasks. (tracking and fuel management). There were no significant differences in performance on the resource management task or the tracking task between the reliability groups. Therefore, it can be concluded that the decrease in monitoring, as a function of automation reliability, was not due to the trading of resources to accommodate higher performance on the manual tasks.

These results partially support previous findings by May et al., (1993) indicating that the more reliable an automated system, the poorer the operator's monitoring performance will be but only to a certain level. The decrease in the detection rate of automation failures with increasing reliability was not completely linear. These findings indicate that poor automation monitoring reflects over-reliance on automation as reliability increases.

REFERENCES

Chambers, N., & Nagel, D.C. (1985). Pilots of the future: Human or computer? Communications of the Association for Computing Machinery, 28, 1187-1199.

Comstock, J. R. & Arnegard, R. J. (1990). Multi-attribute task battery. (Preliminary Technical Report). Hampton, VA: NASA Langley Research Center.

Davies, D.R., & Parasuraman, R. (1982). The psychology of vigilance. London: Wiley.

Lee, J. & Moray, N. (1992). Trust, control strategies, and allocation of function in the human-machine systems, Ergonomics, 35, 1243-1270.

May, P.A., Mollory, R.J., & Parasuraman, R. (1993). Effects of automation reliability and failure rate on monitoring performance in a multi-task environment. Poster session at the 37th Annual Meeting of the Human Factors and Ergonomics Society. Santa Monica, CA: Human Factors and Ergonomics Society.

Mosier, K., Skitka, L.J., & Korte, K.J. (1994). Cognitive and social psychological issues in flight crew/automation interaction. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 191-197). Hillsdale, NJ: Erlbaum.

Mouloua, M., & Parasuraman, R. (1994). Human performance in automated systems: Recent research and trends. Hilsdale, NJ: Erlbaum.

Mouloua, M., & Koonce, J., (1997). Human-interaction: Research and practice. Hillsdale, NJ: Erlbaum.

Muir, B. (1988). Trust between humans and machines, and the design of decision aides. In E. Holnagel, G. Mancini, & D. Woods (Eds.), Cognitive Engineering in complex dynamic worlds. London: Academic Press.

Muir, B. (1989). Operator's trust in and use of automatic controllers in a supervisory process control task. Doctoral thesis. University of Toronto.

Parasuraman, R. & Bowers, J. C. (1987). Attention and vigilance in human-computer interactions. In A. Gale & B. Christie (Eds.), Psychophysiology and the Electronic Workplace. London: Wiley.

Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance consequences of automation-induced "complacency". The International Journal of Aviation Psychology, 3, 1-13.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. Human Factors, 28, 665-679.

Scerbo, M., & Mouloua, M. (1999). Automation technology and human performance: Current research and trends. Hillsdale, NJ: Erlbaum.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation induced "complacency": Development of a Complacency Potential Rating Scale. The International Journal of Aviation Psychology, 3, 111-121.

Wiener, E.L. (1988). Cockpit automation. In E.L. Wiener & D.C. Nagel (Eds.), Human factors in aviation (pp. 433-461). San Diego: Academic.

THE ROLE OF CHOICE AND INDIVIDUAL DIFFERENCES IN THE MITIGATION OF NOISE AND TASK STRESSORS ON VIGILANCE

H.C. Neil Ganey[1], James L. Szalma[1,2], Peter A. Hancock[1,2], Mustapha Mouloua[1], Anna Connelly[1], Joseph Dalton[1], and Fleet Davis[1]

[1]MURI-OPUS Laboratory
[2]Institute for Simulation and Training
University of Central Florida
Orlando, FL

The present study investigated the effects of perceived control over aspects of a task (difficulty) on stress and workload in a vigilance task. Additionally, individual differences in dispositional factors and coping strategy chosen by the operator were considered. Prior research has indicated that pessimism has an effect on the level of experienced stress. Additionally, coping strategies have been linked to stress and workload ratings in a variety of tasks (Matthews & Campbell, 1998). While these factors have been investigated independently, as is frequently the case, their complex interaction has remained unaddressed. In the study, we found that optimism was negatively related to task engagement, though it was positively related to distress. Additionally, it was found to be a predictor of task-focused coping behaviors. More conscientious people were less likely to engage in task-focused or emotion-focused coping. Those who were less pessimistic tended to have heightened worry over the task.

INTRODUCTION

Individual differences in operator characteristics influence vigilance performance and the stress experienced by people engaged in such tasks (Berch & Kanter, 1984; Helton, Dember, Warm & Matthews, 1999). Some of those differences, such as personality traits and response strategy, have been shown to influence perceived workload on a variety of tasks (Damos, 1988). One of the goals for the present investigation was to extend the findings of Szalma (2002) on the impact of dispositional pessimism, and differences in the coping strategies observers employ, on stress and workload of vigilance. In this study these effects were extended to included noise stress and the impact of perceived control on the workload of stress and vigilance. Following Dember, Galinsky, and Warm (1992), perceived control was manipulated by providing participants the illusion that they could choose the difficulty level of the vigil. Dember and his colleagues observed that such a manipulation attenuated the vigilance decrement, and it is possible that perceptions of control may also reduce the workload and stress of vigilance.

Pessimism and Performance

In addition to predicting stress-state, Szalma (2002) also observed an impact of pessimism on coping strategies adopted by observers. As Matthews and Campbell (1998) have noted, the study of coping strategies for dealing with workload and stress has received little attention in human factors. To address this gap, they developed a coping inventory for task stressors (CITS) for immediate post-task assessment of coping strategies that individuals use. Three strategies were identified based upon Lazarus and Folkman's (1984) transactional theory of stress. These are task-focused coping, in which the individual copes with stress through behavior aimed at changing the stressful situation, emotion-focused coping, which is directed at changing one's emotional response, and avoidance-focused coping, in which the individual diverts attention away from the task. Individuals high in pessimism tend to engage in emotion-focused coping, while optimists tend to use task-focused coping (Scheier & Carver, 1987). Moreover, while Helton and his colleagues (1999) did not directly measure coping strategy, their findings that pessimists reported higher levels of distress and less task engagement suggest that pessimists were employing emotion-focused coping, since post-task distress implies that form of coping, and optimists were more likely to use task-focused coping, since higher task engagement implies use of that strategy (Matthews & Campbell, 1998). Indeed, Szalma (2002) observed that pessimists were more likely to engage in emotion-focused coping, although that variable did not predict the degree of task or avoidant coping adopted by operators.

Stress State and Performance

Matthews and Campbell (1998) observed consistent patterns of correlation among coping strategies and the post-task stress states of observers, as measured by the Dundee Stress State Questionnaire (DSSQ), a multi-dimensional instrument for assessing transient states associated with mood, arousal, and fatigue (Matthews, et al., 1999). The DSSQ consists of 10 factor analytically determined scales, which load onto three secondary factors that reflect the individual's stress state. Three of the scales reflect Task-Engagement (energetic arousal, motivation, and concentration), three scales

constitute the dimension of Distress (tense arousal, confidence and control, and hedonic tone), and four scales define the dimension of Worry (self-focused attention, self-esteem, task relevant cognitive interference, and task-irrelevant cognitive interference). Matthews and Campbell (1998) observed that task-coping correlated with energetic arousal and motivation, indicating greater task-engagement, while emotion-focused and avoidance-focused coping were both related to increased worry, reflected by both kinds of cognitive interference. Emotion-focused coping was also linked to higher tense arousal and lower hedonic tone (increases in distress), while avoidance-focused coping was linked with a drop in energetic arousal (lower task-engagement). In view of these findings, a goal for this study was to investigate how differences in pessimism and optimism impact the stress associated with vigilance tasks. The main hypotheses were that:

- Those who experience noise conditions will be more distressed and less engaged than those who do not, but the effect will be attenuated by having perceived control.
- Pessimists will show increased distress and lower task engagement than optimists.
- Pessimists will be more likely to engage in emotion-focused coping, while optimists will be more likely to engage in task-focused coping.

METHOD

Participants

Forty-four undergraduate Psychology students (22 female, 22 male) from a southeastern university participated in the study for extra credit. The students ranged in age from 18 to 45 with a mean age of 23.21. All participants had normal or corrected-to-normal vision.

Instruments

Adjective Checklist.. The adjective checklist is based on of the Big Five Factor Theory of Personality. The participants are asked to respond to how representative they feel the adjectives are to themselves. The five constructs measured are extraversion, conscientiousness, agreeableness, neuroticism, and openness to experience.

DSSQ. The Dundee Stress State Questionnaire (DSSQ) (Matthews et al., 1999) is a 104-item multidimensional self-report instrument for assessing transient states. Specifically, the DSSQ is made up of eleven sub-scales measuring Energetic Arousal, Tense Arousal, Hedonic Tone, Motivation, Self-focused Attention, Self-Esteem, Concentration, Confidence and Control, Task Relevant Interference, and Task Irrelevant Interference. These ten scales load onto three secondary scales: distress, worry, and task engagement.

OPI. The Optimism-Pessimism Index (Dember et al., 1989) is a fifty-six-item questionnaire (eighteen items indicating optimism (O), eighteen items indicating pessimism (P), and twenty filler items) that directly taps expectations. Respondents are asked to rate their agreement with the items using a four-point Likert-type scale ranging from 1 (strongly agree) to 4 (strongly disagree). Unlike other measures of optimism-pessimism (such as the ASQ and LOT), the O/P measures pessimistic tendencies separately from optimistic tendencies. In the initial study with this instrument (Dember et al., 1989), the items on the O scale had a coefficient alpha of .84. The items on the P scale had a coefficient alpha of .86. In another study (Dember and Brooks, 1989), O had a test-retest reliability correlation of .75 over a two-week interval, and P a test-retest reliability correlation of .84.

Procedure

Upon arrival, participants were given an informed consent form, which they read and signed before proceeding. Following the informed consent, participants took the OPI, the Adjective checklist, and the pre-task DSSQ. They worked through a 15-minute practice session of the task, which gave them response-based feedback of hit, miss, or false alarm. No feedback was given for correct rejections. The twenty participants were then assigned at random to one of the four conditions: control-noise, control-silent, choice-noise, and choice-silent. Choice was manipulated by asking the participant if they would like an easy task or a hard task. This was followed by a 30-minute vigilance task, based on one of the focused attention tasks from Szalma's (2002) study, which was divided into four continuous 7.5-min periods of watch. Immediately following the task, participants completed the post-task version of the DSSQ.

Noise

In the present study, noise was used as an environmental stressor in half of the conditions. For this experiment, noise was presented in 1 second bursts of white noise that were presented through a Creative Labs Audigy sound card to Sennheiser HD270 headphones at 90 dB. The gaps between noises were randomized, ranging from 10 seconds to 75 seconds.

RESULTS

In addition to the Individual Differences variables discussed earlier, the impact of two experimental manipulations (noise and choice) was included in the regression analyses. The influence of personality, noise, and choice was evaluated via regression, using dummy coding for the categorical variables.

Noise and Choice

With regard to the state-based stress measures, the only dimension noise influenced was change in distress from pre-task to post-task. An interaction between noise and pessimism was found to predict distress, $F(3,40)=6.291$, $p=.024$. An additional interaction between choice and optimism was also found to predict distress, $F(3,40)=6.618$, $p=.019$. Separate regression analyses for noise and no noise conditions of the data, seen in Table 1, revealed that when noise was absent, those people who were pessimistic were less distressed. This difference is absent when noise is present. Separate regression

analyses for choice and no choice conditions, seen in Table 2, showed that those high in optimism tended to be more distressed when they did not have choice, though this effect is eliminated when they are given choice.

Table 1. Summary of Regression Analysis for Pessimism predicting Distress, for each Noise condition.

	B	Std. Error	β	Sig.
P (no noise)	-.141	.029	-.736	<.0005*
P (noise)	-.014	.056	-.056	.805

Note: * p<.05 # .1>p>.05

Table 2. Summary of Regression Analysis for Optimism predicting Distress, for each Choice condition.

	B	Std. Error	β	Sig.
O (no choice)	.176	.064	.526	.012*
O (choice)	.061	.040	.328	.136

Note: * p<.05 # .1>p>.05

Pre- and Post-Task Stress State

Results of regression analyses of each pre-task stress state on optimism and pessimism are shown in Tables 3-5. Supporting the findings of Szalma (2002), optimism significantly predicted both task engagement and distress before the task, with those high in optimism feeling less pre-task distress and more pre-task engagement. Pessimism, however, did not predict either pre-task distress or task engagement, as it had in Szalma's (2002) work. Pessimism did predict pre-task worry, with those higher in pessimism feeling more pre-task worry. Neither optimism nor pessimism was found to predict any of the post-task stress states significantly.

Table 3. Summary of Regression Analysis for Variables Predicting Pre-Task Engagement

	B	Std. Error	β	Sig.
Optimism	.047	.017	.446	.009*
Pessimism	-.006	.016	-.062	.704

Note: * p<.05 # .1>p>.05

Table 4. Summary of Regression Analysis for Variables Predicting Pre-Task Distress

	B	Std. Error	β	Sig.
Optimism	-.106	.026	-.558	<.0005*
Pessimism	.030	.024	.176	.209

Note: * p<.05 # .1>p>.05

Table 5. Summary of Regression Analysis for Variables Predicting Pre-Task Worry

	B	Std. Error	β	Sig.
Optimism	.013	.027	.075	.633
Pessimism	.094	.025	.586	.001*

Note: * p<.05 # .1>p>.05

Coping

Results for regression analyses of each coping strategy on personality measures and pre-vigil stress state are shown in Tables 6-8. Optimism significantly predicted the use of task-focused coping, with those who scored higher in optimism being more likely to utilize task-focused coping strategies.

Pre-task engagement significantly predicted avoidant focused coping, with those who were less engaged being more likely to employ avoidant coping strategies.

Conscientiousness was a significant predictor of both task-focused and emotion-focused strategies, with those scoring higher in conscientiousness being less likely to use either task or emotion-focused coping strategies.

Table 6. Summary of Regression Analysis for Variables Predicting Task-focused Coping.

	B	Std. Error	β	Sig.
Optimism	.421	.203	.419	.046*
Pessimism	.050	.172	.055	.770
Conscientiousness	-.088	.037	-.336	.024*
Extraversion	-.002	.031	-.011	.938
Neuroticism	.001	.046	.004	.982
Pre-Distress	-.988	1.052	-.186	.354
Pre-Engagement	2.055	1.423	.218	.158
Pre-Worry	1.268	.932	.222	.182

Note: * p<.05 # .1>p>.05

Table 7. Summary of Regression Analysis for Variables Predicting Emotion-focused Coping.

	B	Std. Error	β	Sig.
Optimism	.029	.144	.044	.843
Pessimism	-.029	.121	-.050	.810
Conscientiousness	-.068	.026	-.398	.015*
Extraversion	-.005	.022	.032	.835
Neuroticism	.044	.032	.237	.186
Pre-Distress	.412	.745	.119	.584
Pre-Engagement	-.784	1.008	-.128	.442
Pre-Worry	1.220	.660	.328	.073

Note: * p<.05 # .1>p>.05

Table 8. Summary of Regression Analysis for Variables Predicting Avoidance-focused Coping.

	B	Std. Error	β	Sig.
Optimism	.160	.175	.188	.366
Pessimism	-.127	.147	-.164	.396
Conscientiousness	-.051	.032	-.231	.120
Extraversion	-.020	.027	-.102	.473
Neuroticism	.062	.039	.256	.126
Pre-Distress	.036	.904	.008	.969
Pre-Engagement	-4.660	1.223	-.582	.001*
Pre-Worry	1.519	.801	.314	.066

Note: * p<.05 # .1>p>.05

DISCUSSION

Our findings for optimism supported those of Szalma (2002), with optimism being positively related to pre-task distress and task engagement. Additionally, it is of note that

when optimists are given choice, the increases in distress that accompanied the task are missing.

Note that pessimism did not predict emotion-focused coping, while optimism did not significantly predict task-focused coping. While this was not expected, based on the general literature on pessimism, it is consistent with the results obtained by Szalma (2002), suggesting that the relation between disposition and coping may depend on the task domain.

The collective results of this experiment and other studies examining the effects of pessimism on stress and coping indicate that designing vigilance tasks for individuals who differ in their characteristics may depend on the nature of the display to be monitored. Further research will investigate the task parameters that influence the relation between optimism/pessimism and stress and coping. In addition, future research will address performance differences due to noise and choice, and workload effects associated with the choice and noise conditions will be evaluated.

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

REFERENCES

Berch, D., & Kanter, D. (1984). Individual differences. In J. Warm (Ed.), *Sustained attention in human performance* (pp. 143-178). Chichester: Wiley.

Damos, D. (1988). Individual differences in subjective estimates of workload. In P. Hancock (Ed.), *Human Mental Workload* (pp. 231-237). Amsterdam: North-Holland.

Dember, W., Martin, S., Hummer, M., Howe, S., & Melton, R. (1989). The measurement of optimism and pessimism. *Current Psychology: Research and Reviews, 8,* 102-119.

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Loaad Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North-Holland.

Helton, W.S., Dember, W.N., Warm, J.S., & Matthews, G. (1999). Optimism, pessimism, and false failure feedback: Effects on vigilance performance. *Current Psychology, 18,* 311-325.

Lazarus, R., & Folkman, S. (1984). *Stress, appraisal, and coping.* New York: Springer-Verlag.

Matthews, G., & Campbell, S. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society, 42,* 821-825.

Matthews, G., Campbell, S.E., & Falconer, S. (2001). Assessment of motivational states in performance environments. *Proceedings of the Human Factors and Ergonomics Society, 45,* 906-910.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (v.7), (pp.335-350). Tilburg: Tilburg University Press.

Pedhazur, E.J. (1997). *Multiple regression in behavioral research,* 3rd ed. Ft. Worth: Harcourt Brace.

Scheier, M.F., & Carver, C.S. (1987). Dispositional optimism and physical well being: the influence of generalized outcome expectancies on health. *Journal of Personality, 55,* 170-210.

Szalma, J.. Individual differences in the stress and workload of sustained attention. *Proceedings of the Human Factors and Ergonomics Society, 46,* 1002-1006.

Individual Differences in Dispositional Pessimism, Stress, and Coping as a Function of Task Type

Jennifer E. Thropp, James L. Szalma, Jennifer M. Ross, and Peter A. Hancock

Psychology Department and
Institute for Simulation and Training
University of Central Florida
Orlando, FL

Individual differences in dispositional pessimism and choice of coping strategy on performance and stress, in target detection were investigated. The results were consistent with prior research indicating that higher levels of pessimism were associated with higher levels of stress and less effective coping strategies. Similarly, pessimism predicted emotion focused coping only in tasks with spatial uncertainty. There was evidence that the influence of personality on post-task stress may be mediated by pre-task state. Pessimism also predicted avoidant coping and task focused coping, although this prediction was only meaningful in the context of the combination of a temporal discrimination and spatial uncertainty. Hence, the degree to which pessimistic operators engage in the three coping strategies depends upon the characteristics of the task.

One way in which stress impacts performance on perceptual tasks is through perceptual distortions resulting from a 'narrowing' of the range of cues individuals extract from the surrounding environment. The impetus for research in this area was provided by Easterbrook (1959), who asserted that emotional arousal restricts the utilization of the range of cues from the sensory environment. Easterbrook's proposal is spatial in conception, implying that under conditions of chronic or acute stress, peripheral stimuli are sacrificed at the expense of more centrally located cues. As attention narrows (i.e. as the number of cues attended to is reduced), performance capacity is preserved by the retention of focus on salient cues. Eventually, performance fails as stress increases and even salient cues become excluded. However, early investigations aimed at testing Easterbrook's model of attentional narrowing under stress confounded sensory and attentional effects, largely because of an emphasis on the visual modality. More recent evidence indicates that 'narrowing' is an attentional effect and that cue salience can dictate the locus in the sensory field toward which such limited attentional resources are focused (see Dirkin & Hancock, 1984). Narrowing extends beyond the visual sensory domain, and in this paper we are particularly concerned with the common mechanism(s) controlling the perception (and distortion) of space and time.

Hancock and Weaver (in press) proposed that the narrowing that occurs on the spatial dimension also occurs in perception of time, and that these distortions have a common mechanism. This assertion implies that perceptual distortions of space and time under stress covary, and that these processes draw on a common attentional resource pool. An initial test of this hypothesis is reported in a companion paper in these proceedings (Ross, Szalma, Thropp, & Hancock, 2003).

This paper is concerned with how individual differences in personality impact performance and stress. If perception of space and time covary under stress, it is possible that operator characteristics may influence these effects. Indeed, the transactional model of stress (Lazarus & Folkman, 1984; Matthews, 2001) emphasizes the importance of the interaction between the person and his/her environment in determining stress states. Specifically, an individual's traits may exert their influence on stress responses via influences of appraisal of perceptions of the task and distortions that result from those appraisals. If temporal and spatial processing draw upon common resources, then the combining these components should debilitate performance and increase stress to a greater degree in individuals vulnerable to such effects. Further, the way in which individuals cope with stress may vary as a function personality. Drawing on the Lazaraus and Folkman model, Matthews and Campbell (1998) describe three primary forms of coping with stress associated with task performance; these are task-focused coping, emotion-focused coping, and avoidant coping. In task-focused coping, the individual focuses on dealing with the task and develops strategies aimed at accomplishing the goals of the task. Emotion-focused coping emphasizes the individual's feelings regarding the task; this may motivate a desire to change these feelings as needed. Additionally, Scheier and Carver (1987) indicate that emotion-focused coping is used when the individual does not feel in control of alleviating the stressor, concluding that nothing can be done to ameliorate it. The third coping strategy, avoidant coping, involves diverting one's attention from the stressful situation (see also Cox & Ferguson, 1991).

Pessimism is a characteristic that has been shown to be linked with vulnerability to stress, and it has been well established that pessimists tend to cope poorly with stress (e.g., Dember, Martin, Hummer, Howe, & Melton, 1989; Scheier & Carver, 1987). Thus, Szalma (2002) observed that pessimism predicted increased stress and predicted emotion-focused coping in a vigilance task. Helton, Dember, Warm,

and Matthews (1999) reported that pessimists exhibited steeper vigilance decrements and were more distressed and less task engaged after a vigil than prior to its start. It is likely that the impairments in performance and the increased stress experienced by individuals high in pessimism are due in part to the consumption of resources by maladaptive coping responses. Hence, if combining spatial and temporal components of a task consumes more resources than tasks which only require processing primarily along one of these dimensions, pessimists may report higher levels of stress and may perform more poorly as they devote more resources to cope with the increased demands presented by the task. These effects should be exacerbated in the presence of noise, which pessimists should respond to with greater levels of stress and maladaptive coping strategies. In the companion article published in these proceedings, Ross et al. (2003) reported that intermittent white noise increased the stress associated with task performance. It may be that these effects will be stronger in more pessimistic individuals, as noise presents another form of stress with which observers must cope.

METHOD

Forty-four students, 22 females and 22 males enrolled in psychology classes at the University of Central Florida, served as participants in order to receive extra credit applicable to their course(s). They ranged in age from 18 to 32 years, with a mean age of 21.3 years. All participants engaged in three tasks consisting of a spatial-only task, a temporal-only task, and a spatial-temporal combination task. The order in which the tasks were completed was counterbalanced across participants to control for potential order effects. Eleven females and 11 males were assigned at random to each of the stress conditions (noise and quiet).

Pessimism was measured using the Optimism/ Pessimism Inventory (OPI: Dember, Martin, Hummer, Howe, & Melton, 1989). The instrument was administered prior to receiving instructions for the first task.

All participants engaged in three 7-minute target detection tasks (spatial, temporal, and spatial-temporal combination) during which they monitored the repetitive presentation of gray capital letters consisting of O, D, and a backwards D on a video display terminal (VDT) presented against a visual mask consisting of white unfilled circles on a black background. In each task the event rate was 24 events per minute.

In the spatial-only task letters appeared on the screen in any one of nine prescribed locations. This location variation provided a degree of spatial uncertainty. Participants were required to detect an O of higher luminance. Distracters consisted of an O of lower luminance or a D or backwards D of either luminance. Each letter appeared on the screen for 500 ms, and the order and location of the critical signals and neutral events was randomized across the session.

In the temporal-only task letters appeared on the display in the center of the screen in every trial, thereby precluding any spatial uncertainty. In this task observers were required to detect an O of shorter duration (300 ms), and the luminance of the letters was held constant. Non-targets were an O of longer duration (500 ms) or a D or backwards D of either duration. The order of the critical signals and neutral events was randomized across the session.

In the spatial-temporal combination task, letters appeared on the screen in one of the same prescribed locations as utilized in the spatial task. In this task, the temporal discrimination was combined with spatial uncertainty regarding target location. The order and location of the critical and neutral signals was randomized across the session.

Participants in the noise condition experienced 85dB intermittent white noise presented through headphones at random times during each session. Four noise durations were used: 4, 6, 8, and 10 seconds. Each duration was presented ten times during each task session. Participants in the quiet condition also wore headphones to control for potential testing condition effects and to filter ambient noise.

Pre- and post-task stress was measured using the Dundee Stress State Questionnaire (DSSQ), a multi-dimensional instrument for assessing transient states associated with mood, arousal, and fatigue (Matthews et al., 1999). The DSSQ consists of eleven factor-analytically determined scales which load onto three secondary factors that reflect the individual's stress state. The scales consist of the following: Task Engagement (composed of Energetic Arousal, Concentration, and two motivation scales, Intrinsic and Success), Distress (composed of Tense Arousal, Hedonic Tone, and Confidence and Control scales), and Worry (consisting of Self-Focused Attention, Self-Esteem, Task Relevant Cognitive Interference, and Task Irrelevant Cognitive Interference). Note that the post-task version of the DSSQ includes the Coping Inventory for Task Stress (Matthews & Campbell, 1998), which provides measures for task-focused, emotion-focused, and avoidant coping responses to task stress.

For the spatial and combination tasks, participants responded to targets by pressing the key in the numerical keypad which corresponded to the location of the screen in which the critical signal appeared. For the temporal task, participants responded to critical signals by pressing the 5 key in the numerical keypad because all signals appeared only in the center position of the screen. Participants were asked to remove their watches and turn off the ringers of their cell phones and beepers to minimize acoustic distractions.

RESULTS

Preliminary analyses indicated that there were no significant interactions between pessimism and noise for performance (sensitivity and response bias) or stress. However, there was a significant main effect for pessimism

predicting general pre-task distress levels, $F(1, 42) = 8.46$, $p < .01$, indicating that higher pessimism produced higher distress prior to task commencement. In addition, there was a main effect for pessimism predicting general pre-task worry, $F(1, 42) = 15.49$, $p < .0005$, such that increased pessimism also resulted in greater worry prior to executing the tasks.

The remaining data were analyzed with regression equations to predict the task-specific influences of individual differences. The results of these regressions are displayed in Tables 1-3.

Spatial Task

Pessimism predicted avoidant coping, $F(1, 42) = 12.71$, $p < .005$, post-task engagement with pre-task engagement entered in the model, $F(2, 41) = 10.78$, $p < .0005$, and post-task distress, $F(1, 42) = 8.70$, $p < .01$. Pessimism did not predict task-focused coping, emotion-focused coping, or post-task engagement in the temporal task ($p > .05$). Individuals with greater pessimism in the spatial task demonstrated significantly more avoidant coping, less post-task engagement (with pre-task engagement entered in the model), and also exhibited more distress after completing the task. Additionally, there was a trend for an interaction between pessimism and noise to predict emotion-focused coping (with pessimism and noise entered in the model, $p_{\Delta F} > .05$), where individuals with increased pessimism in the noise condition tended to engage in less emotion-focused coping.

Temporal Task

In the temporal task, pessimism significantly predicted avoidant coping, $F(1, 42) = 6.84$, $p < .05$, and post-task distress, $F(1, 42) = 4.63$, $p < .05$. Pre-task distress significantly predicted post-task distress, $F(1, 42) = 29.55$, $p < .0005$, yet, adding pessimism did not account for any additional variance in post-task distress ($p_{\Delta F} > .05$). Pessimism did not significantly predict emotion or task focused coping or post-task engagement in the temporal task ($p > .05$). Thus, increased pessimism in the temporal task was associated with the tendency to engage in significantly more avoidant coping and to also experience more stress before beginning the task. In addition, high levels of pre-task distress also indicated high levels of post-task distress.

Spatial-temporal Combination task

Pessimism significantly predicted avoidant coping, $F(1, 42) = 23.85$, $p < .0005$, task-focused coping, $F(1, 42) = 11.26$, $p < .0005$, emotion focused coping, $F(1, 42) = 8.48$, $p < .01$, post-task engagement, with pre-task engagement entered in the model, $F(1, 42) = 14.46$, $p < .0005$, and post-task distress, $F(1, 42) = 7.43$, $p < .01$. Thus, increased pessimism in the spatial-temporal combination task indicated significantly more avoidant coping, less task-focused coping, less post-task engagement (with pre-task engagement in the model), and more distress after completing the task.

Table 1.
Summary of Multiple Regression Analysis for Pessimism as a Predictor of Coping Strategies (N = 44)

Task	*B*	*SE B*	*β*	R^2
Spatial Task				
Task Coping	-0.11	0.11	-.16	.03
Emotion Coping	0.22	0.09	0.37	.13*
Avoidant Coping	0.47	0.11	.55	.30*
Temporal Task				
Task Coping	0.09	0.11	-.13	.02
Emotion Coping	0.1	0.07	0.2	.04
Avoidant Coping	0.24	0.09	.37	.14**
Combination Task				
Task Coping	-0.34	0.10	-.46	.21*
Emotion Coping	0.2	0.07	0.41	.17**
Avoidant Coping	0.47	0.10	.60	.36**

Note. * $p < .05$: ** $p < .01$

Table 2.
Summary of Multiple Regression Analysis for Pessimism as a Predictor of Post-Task Distress and Post-Task Engagement, with Appropriate Pre-Task States Entered in Model (N = 44)

Task	*B*	*SE B*	*β*	R^2
Spatial Task				
Post-task Distress	0.02	0.11	.16	.49
Post-Task Engagement	-0.04	0.02	-.33	.45*
Temporal Task				
Post-task Distress	-0.009	0.02	.06	.42
Post-Task Engagement	0.008	0.01	-.08	.48
Combination Task				
Post-task Distress	0.02	0.02	.17	.39
Post-task Engagement	-0.05	0.02	-.36	.55**

Note. *$p < .01$; **$p < .005$

Table 3.
Summary of Multiple Regression Analysis for Pessimism as a Predictor of Post-Task Distress (N = 44)

Task	*B*	*SE B*	*β*	R^2
Spatial	0.06	0.02	.41	.17*
Temporal	0.05	0.02	.32	.10**
Combination	0.06	0.02	.39	.15*

Note. * $p < .01$; **$p < .05$

DISCUSSION

A goal for this study was to determine the impact of dispositional pessimism on stress and coping for three different tasks in the context of noise stress. Consistent with Szalma (2002), pessimism predicted emotion-focus coping, confirming that the higher stress and poor coping strategies observed among individuals with greater pessimism in other domains extends to target detection tasks. In addition, higher levels of pessimism predicted greater declines in post-task engagement, but only in tasks with spatial uncertainty. Although pessimism also predicted post-task distress, these effects were attenuated when the pre-task state was entered into the regression model.

Similarly, pessimism predicted emotion focused coping only in tasks with spatial uncertainty. Pessimism predicted less task-focused coping, but only in the combined task. Taken together, these task-specific results regarding engagement and emotion-coping indicate that more pessimistic individuals become disengaged from the task in an attempt to cope (using emotion focused coping) with the spatial uncertainty that confronts them. These results may be a manifestation of attentional narrowing along the spatial dimension.

In contrast to Szalma (2002), however, pessimism also predicted avoidant coping, consistent with the general literature on pessimism (Scheier & Carver, 1987). In addition, unlike Szalma (2002), the results of this study indicate that pessimism predicted less task focused coping, although this prediction was only meaningful in the context of the combination of a temporal discrimination and spatial uncertainty. Hence, the degree to which pessimistic operators engage in the three coping strategies depends upon the characteristics of the task.

Considered with the results of Szalma (2002), the findings of the current study reveal that how individual characteristics moderate stress effects vary as a function of task characteristics. Future research efforts will further explore task characteristics which interact with operator traits in determining the stress of performance in target detection. This includes an experiment comparing temporal and spatial discriminations using the same methodology reported here.

The findings are consistent with Helton et al. (1999) in that individuals with increased pessimism became more emotion-focused and therefore less attentive under stress. This lack of task-focused attention may explain the general tendency for those with pessimistic proclivities to feel less engaged by the task. This tendency is consistent with Scheier and Carver (1987) who state that emotion-focused coping is adopted when individuals feel they can do nothing to solve the problem; perhaps this is why they do not feel particularly engaged. The findings of the present study can be reconciled with Hockey and Hamilton's (1983) account differences in selectivity. The observer who is minimally engrossed by the task and instead focuses on task-irrelevant stimuli may be likened to the pessimistic individual who reports low task engagement and high avoidant coping. Correspondingly, the participant who appears highly attentive to the activities required to complete the task is reminiscent of the optimistic individual who experiences high task engagement.

More pessimistic individuals engaged in less task-focused coping only in the spatial-temporal combined tasks. This tendency to divert one's attention away from the task and instead execute maladaptive coping techniques is an indicator of increased stress, especially in those who are more susceptible to stress (e.g. pessimists). Because this task demanded both spatial and temporal resources and promoted more stress (as seen in the adoption of maladaptive coping), it is possible that these two task components draw upon the same resource pool, thereby increasing the demand placed on the operator. In accordance with the Lazarus and Folkman (1984) transactional theory of stress, the operator may perceive the task as demanding more resources than can be supplied, thereby appraising the task as stressful.

Note that although pessimism was predictive of stress, it did not predict performance, as indexed by signal detection theory measures of d' and c. This may be due to a ceiling effect for performance, since the sensitivity scores were high across all conditions (d'>5; see Ross, Szalma, Thropp, & Hancock, 2003). Current experimental work will determine whether declines in performance are related to pessimism in more difficult discriminations.

The results of the current study confirm that pessimism can influence stress and coping, and that stress is multidimensional in character (Matthews, 2001). However, these effects do not extend to performance or to the presence of intermittent noise. An explanation for the absence of such effects may be found in the Hancock and Warm (1989) model of stress and sustained attention, which the level of behavioral adaptability varies as a function of stress. Specifically, performance is preserved at moderate stress levels as the organism successfully copes with the demands of the environment and the task, but declines rapidly at the extremes of low and high demand. In addition, there are levels of stress effect, such that subjective state (referred to by Hancock and Warm as 'comfort') declines first, followed by behavioral adaptability, with failure of physiological adaptation representing the final catastrophic failure. In the current experiment observers were able to successfully cope with intermittent noise and task demands, but they were unable to preserve their subjective state. Hence, the design and operation of target detection tasks should consider these differences among individuals. While performance may remain stable, the stress and coping evidence reported here provides further evidence that higher levels of pessimism are associated with greater resources being consumed for non-task-related coping efforts. In such individuals, an abrupt elevation in stress such as an acute increase in task demand or engagement in prolonged work may lead such individuals to performance failure earlier than those who are less pessimistic in disposition.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program, P.A. Hancock, *Principal Investig*ator, administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

REFERENCES

Broadbent, D. E. (1963). Differences and interactions between stresses. *Quarterly Journal of Experimental Psychology,* **15**, 205-211.

Dember, W. N. & Brooks, J. (1989). A new instrument for measuring optimism and pessimism: Test-retest reliability and relations with happiness and religious commitment. *Bulletin of the Psychonomic Society,* **27(4)**, *365-366.*

Dember, W. N., Martin, S. H., Hummer, M. K., Howe, S. R., & Melton, R. S. (1989). The measurement of optimism and pessimism. *Current psychology: Research and reviews,* **8**, 102-119.

Dirkin, G. R. & Hancock, P. A. (1984). Attentional narrowing to the visual periphery under temporal and acoustic stress. *Aviation, Space, and Environmental Medicine,* **55**, 457.

Easterbrook J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review,* **66**, 183-201.

Hamilton, P., Hockey, G. R. J., & Rejman, M. (1977). The place of the concept of activation in human information processing: An integrative approach. In S. Dornic (Ed.), *Attention and Performance,* Vol. 6. Hillsdale, N. J.: Lawrence-Erlbaum.

Hancock, P. A. & Warm, J. S. (1989) A dynamic model of stress and sustained attention. *Human Factors,* **31(5)**, 519-537.

Hancock, P. A. & Weaver, J. L. (in press). On distortion under time stress. *Theoretical Issues in Ergonomic Science.*

Helton, W. S., Dember, W. N., Warm, J. S., & Matthews, G. (1999). Optimism, pessimism, and false failure feedback: Effects on vigilance performance. *Current Psychology,* **18**, 311-325.

Hockey. R. & Hamilton, P. (1983). The cognitive patterning of stress states. In G. R. J. Hockey (Ed.), Stress and fatigue in human performance. New York: Wiley.

Lazarus, R. S. & Folkman, S. (1984). Stress, appraisal, and coping. New York: Springer-Verlag.

Matthews, G. (2001). Levels of transaction: A cognitive science framework for operator stress. In P.A. Hancock and P.A. Desmond (Eds.), Stress, workload, and fatigue. (pp. 5-33). Erlbaum: Mahwah, NJ.

Matthews, G. & Campbell, S. E. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society, 42, 821-825.*

Navon, D. & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86,* 214-255.

Norem, J. C. & Cantor, N. (1986). Anticipatory and post-hoc cushioning strategies: Optimism and defensive pessimism in "risky" situations. *Cognitive Therapy and Research, 10(3),* 347-362.

Norman, D. & Bobrow, D. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44-64.

Ross, J. M., Szalma, J. L., Thropp, J. E., & Hancock, P. A. (2003). Performance, workload, and stress correlates of temporal and spatial task demands. ***Proceedings of the Human Factors and Ergonomics Society*, 47**, in press.

Scheier, M. F. & Carver, C. S. (1987). Dispositional optimism and physical well-being: the influence of generalized outcome expectancies on health. *Journal of Personality,* **55**, 170-210.

Szalma (2002). Individual differences in the stress and workload of sustained attention. *Proceedings of the Human Factors and Ergonomics Society, 46, 1002-1006.*

Wickens, C. D. (1984). Processing resources in attention. In R. Parasurman and D. R. Davies (Eds.), Varieties of Attention (pp. 63-104). Orlando, FL: Academic.

AN AFFORDABLE MEDIUM-FIDELITY VIRTUAL ENVIRONMENT SOLDIER SIMULATOR DESIGNED FOR INFANTRY COMBAT RESEARCH

Adams Greenwood-Ericksen, Jeanne L. Weaver, Kelly Burke, Kristy A. Bradley, and Peter A. Hancock
University of Central Florida
Orlando, Florida

Due to the expense of obtaining access to purpose-built virtual environments, it has traditionally been difficult for some researchers to examine questions of interest involving infantry combat performance. As part of on going research into stress and human performance in infantry combat, a highly affordable soldier simulator was constructed using off-the-shelf parts and equipment. Simulators of this type are highly flexible and scalable in design, allowing researchers to fabricate them to the requirements of the study and/or in proportion to available resources. Our demonstration will allow participants to operate the simulator themselves, or observe the demonstrators operating it. Additionally, they will be able to design their own missions and execute them in the simulator, with the assistance of trained personnel. The purpose of the demonstration is to acquaint human factors researchers with a system that demonstrates that solder simulator systems can be developed affordably and easily.

INTRODUCTION

There are many problems facing human factors researchers in explaining the area of infantry combat. Due to the difficulty and expense involved in field studies, the use of simulation provides the most accessible option for researchers examining research problems in this area (Pleban et al, 2002, Driskell & Salas, 1991). While the military, other government agencies, and several private sector companies have invested millions of dollars in dismounted infantry soldier-simulator development, relatively little research has been performed in academic settings using these devices. This dearth of research is likely due, at least in part, to the expense of obtaining access to such systems. Of three systems surveyed during a recent procurement project at the University of Central Florida, the least expensive option examined was projected to cost upwards of $40,000 in hardware alone. This problem has been exacerbated by the tendency of government agencies to develop or contract private companies to develop highly specific, simulator packages, rather than more flexible and general simulations that could be adapted for use by other researchers in a cost effective manner. The result is that many academic institutions are unable to afford the facilities to conduct this kind of context-contingent research. Further, many of the simulators that do exist are designed primarily for training rather than research and lack many of the elements that researchers require for successful experimentation. Moreover, such simulations are commonly very difficult to learn to operate or program, and may therefore require the researcher to obtain the services of a full-time programmer. Therefore, performing infantry combat research using traditional soldier simulators is often too difficult and expensive an undertaking for universities or small laboratories. However, it is possible to build more affordable, simpler substitutes for this type of simulator from easily available and relatively inexpensive "off the shelf" components. The development and examination of the design and capabilities of such a simulator is the purpose of the proposed demonstration.

The present work demonstrates a soldier simulator system assembled for under $6000 from such off-the-shelf components. It should be pointed out that the cost of the simulator could vary considerably from this amount depending on the

needs of the user and the quality of the simulation desired.

DESIGN

System Requirements

The purpose of the present project was to develop a medium-fidelity soldier simulator that would allow the researcher access to large numbers of dependent variables and performance metrics. Other requirements were that the simulator would preserve a high degree of face validity and adhere reasonably well to actual combat tactics and behavior. Additionally, the simulator needed to have good single and multiple simultaneous user capabilities, have an accessible scripting utility, and, most importantly, be affordable for small laboratories and universities.

The Software Package

The solution to these requirements was to adapt existing and readily available hardware and software technologies to fill the roles of purpose built equipment. The heart of any simulator is the software package that powers it. This role was to be filled by the use of a commercially available video game. In order to determine which of the wide variety of games available on the market would be the most appropriate choice, we surveyed and tested approximately a dozen alternatives. This option was available to us in part because of the affordability of this software: at an average price of $48.00, we were able to test many of these programs. Additionally, we were able to obtain some copies less expensively or even cost-free by borrowing copies from individuals at the university.

Ultimately, the game "Tom Clancy's Ghost Recon®" was chosen as the best choice for our software component. While the video game was critically acclaimed, receiving several industry and magazine awards for best PC game of the year, the choice was based largely on the factors which affected its performance as part of a soldier simulator. The simulator has a powerful and relatively accessible mission editor, allowing researchers a high degree of control over the environment and computer controlled actors. It has good face validity and weapons modeling, and enemy and friendly forces that behave in a reasonably realistic manner in the context of the simulation. The dynamics of the game force the user to consider the effects of their own movement, weapon choice, accurate range, and the use of cover and concealment by themselves and their enemies in order to effectively engage the enemy. The game is designed to reward good tactical behavior and punish unrealistic "run and gun" tactics common to first person shooter video games, in which the user races frantically about in the game environment, shooting up the landscape indiscriminately. Thus, game dynamics emphasize the use of proper tactical approaches to problems. Finally, the game is not only designed for single player use, but has well developed multiplayer capabilities for up to 36 networked human operators, who can be physically located as close as the next room or as far as halfway around the globe.

Hardware Components

The physical components of the simulator represent the bulk of the cost of the system, but are also the most flexible. The prototype model used a digital video projector, a projector screen, and a PC equipped with a Pentium 4 with a 2.0 GHz processor, 256 megabytes of RAM, and an nVidia GE Force 4 graphics card with 64 megabytes of RAM. The PC system and the projector chosen largely determine the quality of the simulator. Less expensive models, made from less sophisticated parts, will produce a less graphically impressive environment, while higher quality hardware can boost resolutions up to near DVD quality resolution in 32 bit (True) color. With the best hardware currently available and the highest software settings, the simulator can project nearly photorealistic imagery at almost life size. Sound is another factor in which great variation in quality is possible. For the purposes of maximizing immersion, the current system was designed with very advanced sound capabilities (an Extigy sound card by Audigy and Klipsch 5.1 surround sound speakers). The software component supports full 360 degree surround sound, so this relatively inexpensive feature not only greatly enhances immersion, but supplies task-relevant detail to the

user. Additionally, investing in good quality sound provides good value for its expense, as it is quite important in developing and maintaining immersion (Stanney, 2002).

CAPABILITIES

The system's capabilities are quite impressive for its low cost. Users can perform soldier simulator or virtual environment (VE) based research for a fraction of the cost of purpose built equipment. The scripting and mission design tool is provided with the software and is a menu driven utility that is both powerful and easy to learn. This means that researchers themselves can quickly and easily learn to design their own scenarios without having to rely on a specialist programmer. Through the editor, users can control almost every aspect of the combat environment, from lighting to weather, and can script them to change dynamically throughout the course of the mission. Fog can roll in from the sea, precipitation can be made to fall or cease, or wind can be made to increase or decrease. Explosions or airstrikes can be simulated at certain locations at certain times or in response to player actions. Additionally, the user can control the positions, armament, appearance, disposition, and behavior of enemy and friendly units in as little or as much detail as necessary. This means that the experimenter can control the number of friendly and enemy units, their equipment and armament, their training and quality, and their general behavior in combat. Enemy units can be scripted to panic under fire, use cover or concealment, set ambushes, operate vehicles, and work together effectively (or ineffectively if so scripted) as a team.

Monitoring Events and Data Collection

One of the major strengths of the system is the freedom it allows the user to monitor events in the simulation space. The experimenter can actually enter the mission environment visibly or invisibly to observe events in real time by means of the multiplayer functionality. A replay feature logs all control inputs made by the user and tracks the position and behavior of every other character and element in the simulation. This feature allows the experimenter to replay the mission from multiple angles and multiple viewpoints after the participant has finished the task. Replays can be saved for future review. Additionally, the simulation itself keeps track of several important statistics on a running basis, including number of shots fired, number of hits made, number of kills, shooting accuracy, number of squadmates hit or killed, number of mission objectives completed, in progress, or failed, and the amount of time elapsed since the simulation began. This information is displayed following the completion of every mission in a debriefing report.

Simulator Limitations

The simulator does have some limitations, however. The most glaring of these is the inability to interface with a surrogate weapons system or any elaborate movement control. Since it is designed exclusively as a video game, there is no way to avoid using some kind of PC-based input device. Thus, the simulation must always be based around a keyboard, joystick, mouse, or similar input system. Thus there is no provision available for the simulator to interface with a surrogate weapons system such as many advanced dismounted infantry simulators use (SAST, DISALT, others). Additionally, because of the wide array of behaviors available to the user in the simulator, the interface is by necessity fairly complex, and is somewhat non-intuitive in design due to the limitations of the input devices. We have experimented with alternate input devices and layouts, but the problem still remains difficult to solve without at least 10-15 minutes of subject training. Another shortcoming of the simulator is the limitation of the number of friendly soldiers under the user's control allowed in a single player mission. The game sets an arbitrary limit of six squadmates under the user's direct control in single player mode, divided into up to three fire teams. Additionally, the largest fire team possible is three, so in order to use all six possible squadmates in single player mode, it is necessary to use two independent fire teams. This might pose difficulties in coordination for less experienced users.

In general, however, we think that the present system has great potential for researchers interested in pursuing research questions related to behavioral aspects of infantry combat without the

expenditure of large amounts of money and time. Potential applications for this simulator system in research include navigation, target detection or identification, tests of transfer of training, use of tactical strategies, and tests of short term memory and stress effects.

RESEARCH AGENDA: CURRENT EXEMPLARS

Currently, the University of Central Florida's Multidisciplinary University Research Initiative-Operator Performance Under Stress laboratory is using several versions of this simulator in planned and ongoing research projects included in three proposed courses of study. The simulator is currently being used to test navigation ability using a virtual map and compass in a simulated non-combat search and rescue mission under an 85 dB white noise stressor. A second proposed experiment in this series will address visual spotting and identification of enemies in a simulated embassy defense mission under a similar stressor. Another proposed series of studies is designed to examine naturalistic decision making in a real time combat environment under various stressors, with friendly and enemy forces interacting in real time with the participant and with each other in response to the participant's decisions. The third series of studies builds on work performed using a previous soldier simulator to identify the effects of time stress on performance in a simulated hostage rescue mission.

DEMONSTRATION OF THE SIMULATOR

The proposed demonstration will illustrate the capabilities of the simulation in a hands-on, audience accessible format by allowing observers to interact directly with the simulator and the scripting utility. Participants visiting the demonstration will be able to design and execute their own missions with the assistance of trained UCF graduate research assistants or interact with pre-scripted scenarios. Additionally, the built-in replay feature will be used to show more complex missions being executed and to provide participants in the demonstration with feedback regarding their performance. Additionally, participants in the demonstration will be able to ask questions regarding any aspect of the simulator, from how to build or obtain simulation equipment to questions regarding mission design or infantry combat tactics within the simulation. It is hoped that the present demonstration will provide an opportunity to facilitate conference participants in developing their own research ideas and performing their own experiments using simulators such as that demonstrated. It should be pointed out that the authors are distributing any and all information pertaining to the design and use of the simulator freely; it is hoped that the use of similar systems will encourage research in the vital area of infantry combat performance by making solder simulator technology more widely available to researchers in human factors related fields.

ACKNOWLEDGEMENTS

This research was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

REFERENCES

Driskell, J.E., & Salas, E. (1991). Overcoming the effects of stress on military performance: Human factors, training, and selection strategies. In Gal, R. & Mangelsdorff, A.D. [Eds] *Handbook of Military Psychology*. New York: John Wiley and Sons.

Pleban, R.J., Matthews, M.D., Salter, M., Eakin, D.E. (2002). Training and assessing complex decision making in a virtual environment. *Perceptual and Motor Skills, 94*, 871-882.

Stanney, K. [Ed]. (2002). *Handbook of virtual environments: Design, implementation, and applications.* Lawrence Erlbaum Associates, Publishers, Mahwah, NJ.

COMPARISON OF FUZZY SIGNAL DETECTION AND TRADITIONAL SIGNAL DETECTION THEORY: APPROACHES TO PERFORMANCE MEASUREMENT

Lauren Murphy, James L. Szalma, and Peter A. Hancock

Department of Psychology and Institute of Simulation and Training
University of Central Florida
Orlando, FL.

A recent advance upon Signal Detection Theory (SDT) promises to enhance measurement of performance in complex real world domains where stimuli do not fall into discrete, mutually exclusive categories. This development, Fuzzy Signal Detection Theory (FSDT) combines traditional SDT with Fuzzy Set Theory (FST) to extend signal detection analysis beyond the traditional crisp, categorical decision making model. FSDT allows for events to simultaneously be in more than one state category (e.g. signal and non-signal), so that stimulus and response dimensions can be continuous rather than categorical. This study compared the differences in methods of analyses from FSDT and traditional SDT using the same data set. Data suggests that FSDT analysis and traditional SDT provide different vistas into signal detection performance. FSDT provided a better description of the effects of stimulus uncertainty on observers' response bias and sensitivity. This is because the FSDT model explicitly captures this uncertainty and can provide insight into system performance in domains in which stimulus categories vary along a continuum.

For almost fifty years Signal Detection Theory (SDT) has provided an extremely useful analytical tool for evaluating human and machine performance in both simple and complex systems. As such, it has served as one of the most useful measurement tools and influential theories in psychology (Wickens, 2002). However, even outstanding theories have their limitations, and SDT is no exception. In the SDT model, the state of the world is forced into crisp mutually exclusive categories (i.e., signal versus non-signal; old versus new; etc.), which may not be accurate representations of the true states of the world. To capture the uncertainty inherent in real-world stimuli, the representation of the stimulus dimension must be continuous rather than categorical. Recently, Parasuraman, Masalonis, and Hancock (2000; see also Hancock, Masalonis, & Parasuraman, 2000) have captured this uncertainty by combining elements of SDT with those of Fuzzy Set Theory, in which category membership is not considered mutually exclusive and stimuli can therefore be simultaneously assigned to more than one category. Thus, in FSDT a given stimulus, or more formally, a stimulus selected at random from a distribution of stimuli, may be categorized as both a correct detection and a false alarm depending on the degree to which the stimulus represents a critical event. For instance, a convenient range for a stimulus dimension is one in which the strength of the stimulus varies from 0 (a definite non-signal) and 1 (definite signal), with a signal value of .5 representing maximal uncertainty in the stimulus itself. That is, a stimulus with a signal value of .5 has properties of both signal and non-signal to an equal degree. Implicit in this model is the assumption that signal uncertainty exists not only within the observer (a major insight provided by traditional SDT) but in the stimulus dimension itself. This experiment represents an initial test of FSDT using a perceptual signal detection task. In this study fuzzy stimulus and response dimensions were established (seven categories for each), and difficulty (discrimination) and bias (instructional set) were manipulated to investigate the degree to which the FSDT model is sensitive to these classical SDT manipulations. The FSDT analysis was compared to an analysis using traditional SDT, using methods outlined in MacMillan and Creelman (1991) see also Wickens (2002).

METHOD

Seven students from the University of Central Florida (3 men and 4 women, mean age = 21 years) volunteered to participate in this study.

The stimuli employed in this experiment consisted of two vertically oriented 1.64 cm by

20.32 cm rectangular bars separated by 1.26 cm. Seven different images were presented during the experiment, although participants were not informed of the number of stimulus categories, only that the stimuli would vary between two extremes in color.

In each image one of the two rectangles always remained a checkered pattern of squares consisting of the following two colors in equal proportions; violet squares with RGB (R 204, Gr 0, B 255) and Grey (RGB R 127, Gr 127, B 127). The 2nd vertical rectangle consisted of checkered pattern of violet squares with RGB (R 204, Gr 0, B, 255), and also one of seven saturated versions of yellow squares. Using a variation of a psychophysical function developed by Indow and Stevens (1966) yellow (RGB value; R 255, Gr 255, B 0) was varied in its saturation level in images one through seven for the second vertical rectangle only.

Participants read and signed an informed consent form that explained their rights, received instructions regarding the task and were provided with examples of the stimuli. After instructions, participants completed a 5 min practice session. Both difficulty levels of the task (less difficult vs. more difficult) were presented in the practice session, 2.5 min for each level of difficulty. The order in which the participants received these difficulty levels was counterbalanced. Participants engaged in six, 6-min detection tasks consisting of two difficulty levels (less difficult, stimulus presented with 80 percent RGB value vs. more difficult, stimulus presented with 95 percent RGB value) presented for 125 ms, and three levels of instruction bias (lenient, conservative and neutral). The order in which the participants received the different treatments was counterbalanced using a Latin Square design.

The task required participants to monitor the relative color purity of two vertical lines that would flash on and off the screen and to judge the degree to which the lines were the same or different in 'color'. Participants were instructed to respond to stimuli by rating the degree to which each stimulus was a signal (the rectangles differed in saturation) versus a non-signal (the two bars did not differ in color) by pressing keys 1 through 7, with the response '7' indicating that the stimulus was definitely a signal. No other information regarding the number of possible stimulus categories or the relative frequencies of those categories was provided to the observers. For each condition the event rate was 21 event/min, and each of the 7 different stimuli was presented 18 times during each session. The order in which the stimuli were presented within a condition was randomized.

For the manipulation of response bias, three instructional sets were employed to induce lenient, unbiased, and conservative responding. These instructions were based on those employed by Snodgrass and Corwin (1988). During the task with the conservative instruction set participants were informed that they would receive (10) points for each correct identification, which meant that they correctly identified the degree to which the two lines were the same or different in color. However, they were told that they would lose a (-1) point for each missed signal, which meant that they underestimated the degree to which the two lines were different in color. Finally, they were informed that they would be penalized (-10) points for each false alarm, which meant they overestimated the degree to which the two lines were different in color.

During the neutral (unbiased) condition participants were informed that they would receive a (1) point for each correct identification. However, they were told that they would be penalized a (-1) point for each missed signal and that they would be penalized a (-1) point for each false alarm. For the lenient condition participants were informed that they would receive (10) points for each correct identification. However, they were told that they would be penalized (-10) points for each missed signal and a (-1) point for each false alarm.

RESULTS

For all statistical analyses an alpha level of .05 was adopted, and Box's epsilon was used to correct for violations of sphericity (Myers & Well, 1995). A 2 (difficulty) x 3 (bias instructions) repeated measures analysis of variance was used to test differences among conditions in perceptual sensitivity (d') and response bias (c). Formulas for calculating sensitivity and response bias were the same for both FSDT and Traditional SDT analyses, since the fuzziness is captured in the hit and false alarm rates (Parasuraman et al., 2000). The formulas for calculating FSDT measures were obtained from Parasuraman et al. (2000), and the method for computing hit and false alarm rates in the traditional manner followed procedures outlined by MacMillan and Creelman (1991).

Overall Sensitivity and Response Bias

Mean sensitivity scores for each difficulty level and bias instruction using traditional and fuzzy computation methods are presented in Table 1 and Table 2, respectively. Sensitivity (d') was calculated using the Traditional SDT method of analysis collapsed across stimulus, total d' (see MacMillan & Creelman, 1991, see also Wickens, 2002). An ANOVA revealed that participants were significantly more sensitive in the less difficult conditions compared to the more difficult conditions, $F(1, 6) = 21.12$, $p = .004$. Participants did not significantly differ in their d' scores across bias conditions, ($p>.05$).

An ANOVA of sensitivity scores based on the FSDT method of analysis (collapsed across stimuli) revealed that participants were significantly more sensitive in the less difficult conditions compared to the more difficult conditions, $F(1, 6) = 32.23$, $p = .001$ (see Table 1). In addition there was a significant main effect for bias, $F(2, 12) = 7.97$, $p = .006$ (see Table 2). Data from pairwise comparisons revealed that sensitivity is greater in the conservative condition compared to the neutral condition and lenient.

Table 1

Index of Perceptual Sensitivity (d') collapsed across stimuli

Difficulty level	FSDT Mean (SD)	TSDT Mean (SD)
Less	2.50 (0.10)	13.28 (1.22)
More	2.01 (0.04)	7.42 (1.19)

Table 2

Index of Perceptual Sensitivity (d') collapsed across stimuli

Bias by instruction	FSDT Mean (SD)	TSDT Mean (SD)
Conservative	2.15 (0.52)*	11.27 (0.78)
Neutral	2.31 (0.71)	10.08 (1.05)
Lenient	2.30 (0.95)	9.71 (1.80)

* $p = .006$ for each comparison

When c was calculated using the FSDT method of analysis collapsed across stimuli, the data revealed a significant effect for the response bias manipulation (conservative, neutral, and lenient), $F(2,12)= 8.02$, $p=.02$. The data displayed in Table 3, revealed from the pairwise comparisons, that participants' bias (degree of conservativeness and leniency) in purity discrimination was significantly more conservative for stimuli presented in the conservative condition compared to the participants' bias toward stimuli presented in the lenient and neutral conditions, which did not significantly differ from each other.

Table 3

Index of Response Bias (c) collapsed across stimuli

Bias by instruction	FSDT Mean (SD)	TSDT[a] Mean (SD)
Conservative	0.53 (0.16)*	-
Neutral	0.16 (0.17)	-
Lenient	-0.06 (0.24)	-

[a] An overall (c) cannot be calculated from Traditional Signal Detection Theory because of the issue of summing across multiple response criterions (see MacMillan & Creelman , 1991).

* $p = .02$

Sensitivity and Response Bias Analyzed by Stimulus

When d' was calculated using the FSDT method of analysis by stimulus, an ANOVA revealed that participants' sensitivity was significantly different across stimuli (2 through 6), $F(3, 16) = 7.61$, $p = .03$, and difficulty levels, $F(1, 6) = 12.604$, $p = .01$. Pairwise comparisons indicated that participants' sensitivity for color discrimination was significantly greater for stimuli with maximal uncertainty such as stimulus 4 ($M = 2.47$, $SD = 0.40$), stimulus 5 ($M = 2.35$, $SD = 0.72$), and stimulus 3 ($M = 1.92$, $SD = 0.24$), and lower for stimuli with minimal uncertainty such as stimulus 2 ($M = 1.91$, $SD = 0.35$) and stimulus 6 ($M = 0.74$, $SD = 0.10$). This trend appears to be stronger for the less difficult condition ($M = 2.05$, $SD = 0.22$) compared to the more difficult condition ($M = 1.70$, $SD = 0.17$).

Similarly, when d' was calculated using the Traditional SDT method of analysis by stimulus interval, the analysis revealed that participants were significantly more sensitive in the less difficult conditions ($M = 2.21$, $SD = 0.20$) compared to the more difficult conditions ($M = 1.23$, $SD = 0.20$), $F(1, 6) = 21.12$, $p = .004$. The data revealed a significant main effect for stimuli (i.e., stimuli 1 to 2, 2 to 3, 3 to 4, 4 to 5, 5 to 6, and 6 to 7), $F(5, 30) = 5.26$, $p = .001$, indicating

that participants' sensitivity in color discrimination was significantly different for stimuli intervals. However, in contrast to the findings from the FSDT method of analysis, the traditional analysis indicated that participants discriminated stimulus 6 from 7 (M = 3.64, SD = 0.70) with the highest sensitivity in color discrimination, followed by the interval between stimuli 2 and 3 (M = 2.06, SD = 0.22), stimuli 3 and 4 (M = 1.77, SD = 0.34), stimuli 4 and 5 (M = 1.64, SD = 0.39), and stimuli 5 and 6 (M = 1.20, SD = 0.38). Participants discriminated stimuli 1 and 2 with the least sensitivity (M = 0.03, SD = 0.72).

Response Bias (c) was calculated for each stimulus using the FSDT method of analysis. Note that in FSDT, it is possible to compute a hit and false alarm rate for each stimulus. The exception is the extreme two categories (1 and 7), since it is not possible to correctly detect a signal that is not present (1) or to commit a false alarm when responding to a pure signal (7). An ANONVA revealed a significant effect for stimuli (2 through 6), F (10, 129) = 12.67, p = .007. The data revealed, from the pairwise comparisons, that participants' bias (degree of conservativeness and leniency) in color discrimination was significantly different from stimuli with the least uncertainty (stimuli 2 and 6) compared to the stimuli with the most uncertainty (stimuli 3, 4, and 5). Data revealed that participants responded to stimulus 6 with the highest bias toward a lenient response (M = -1.04, SD = 0.34), followed by stimulus 5 (M =-0.27, SD = 0.42), stimulus 4 (M = 0.139, SD = 0.30), and stimulus 3 (M = 0.56, SD = 0.32). Participants responded to stimulus 2 (M = 1.58, SD = 0.50) with the highest bias toward a conservative response. In contrast, when c was calculated using Traditional SDT method of analysis by stimulus intervals (i.e., intervals 1 and 2, 2 and 3, etc.) revealed no differences where for stimuli.

DISCUSSION

Fuzzy Signal Detection analysis and Traditional Signal Detection provide different vistas into signal detection performance. That FSDT provided a better description of the effects of stimulus uncertainty on observers' response bias and sensitivity is not surprising, given that the FSDT model explicitly captures this uncertainty.

However, the finding that the Traditional SDT model and the FSDT model were perceptually sensitive (d') to different stimuli is interesting. However, it should be recognized that the meaning of d' is different for FSDT model compared to the Traditional SDT model.

The sensitivity index, in the Traditional SDT model, is gauging the participants' ability to place stimuli on the signal dimension, while for FSDT, the index captures sensitivity to the degree of signalness. For FSDT results indicated that participants correctly gauged the degree of signalness and were more sensitive to stimuli in the middle, with the most uncertainty. For Traditional SDT results indicated that participants were most sensitive to the stimulus at the high end of the dimension.

The individual results are not surprising because the FSDT model captures uncertainty in the data and this explains why the data reveals the highest sensitivity for stimuli with the most uncertainty. The results associated with the Traditional SDT analyses conform to the psychometric function of the stimulus dimension. Because the traditional methods are from of multi-dimensional scaling, such a result is not surprising (MacMillan and Creelman, 1991; Wickens, 2002). The surprise is in the very different results that are seen when with the same data set is analyzed using FSDT versus Traditional SDT.

The different findings suggest that further work is necessary to investigate the relation between the index of sensitivity and the FSDT model. These results do confirm, however, that FSDT can be used to analyze system performance in domains in which stimulus categories vary along a continuum. Further, initial evidence suggests that Traditional SDT and FSDT methods may provide different insights into system performance.

ACKNOWLEDGEMENT

This work was facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, Dr. P.A. Hancock, *Principal Investigator*. The assistance of the Department of Psychology at University of Central Florida is also gratefully acknowledged. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

REFERENCES

Hancock, P.A., Masalonis, A.J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science*, **1**, 207-230.

Indow, T. & Stevens, S.S. (1966). Scaling of saturation and hue. *Perception & Psychophysics*, **1**, 253-271.

Macmillan, N.A., & Creelman, C.D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Myers, J.L., & Well, A.D. (1995). *Research Design and Statistical Analysis*. (2nd Ed.). New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors,* **42**, 636-659.

Snodgrass, J.G., & Corwin, D. (1988). Pragmatics of measuring recognition memory. Application to dementia and amnesia. *Journal of Experimental Psychology, General, **117***, 34-50.

Wickens, T.D. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.

Comparison of Fuzzy Signal Detection Theory and Traditional Signal Detection Theory: Approaches to Performance and Measurement

L.L. Murphy, J.L. Szalma, & P.A. Hancock

MIT² INSTITUTE for SIMULATION & TRAINING

Abstract

A recent advance in Signal Detection Theory (SDT) promises to enhance measurement of performance in such complex real world domains. This development, Fuzzy Signal Detection Theory (FSDT; Parasuraman, Masalonis, & Hancock, 2000), combines traditional SDT with Fuzzy Set Theory to extend signal detection analysis beyond the traditional crisp, categorical decision making model. FSDT allows for events to simultaneously be in more than one state category (e.g. signal and non-signal), so that the stimulus and response dimensions can be continuous rather than categorical. This work aims to compare the differences in methods of analyses from FSDT and traditional SDT using the same data set.

Introduction

For almost fifty years Signal Detection Theory (SDT) has provided an extremely useful analytical tool for evaluating human and machine performance in both simple and complex systems. As such, it has served as one of the most useful measurement tool and influential theories in psychology. However, even outstanding theories have limitations, and SDT is no exception. In the SDT model, the state of the world is forced into crisp mutually exclusive categories (i.e., signal versus non-signal; old versus new; etc.), which may not be accurate representations of the true states of the world. To capture the uncertainty inherent in real-world stimuli, the representation of the stimulus dimension must be continuous rather than categorical. Recently, Parasuraman, Masalonis, and Hancock (2000; see also Hancock, Masalonis, & Parasuraman, 2000) have captured this uncertainty by combining elements of SDT with those of Fuzzy Set Theory, in which category membership is not considered mutually exclusive and stimuli can therefore be simultaneously assigned to more than one category. Thus, in FSDT a given stimulus, or more formally, a stimulus selected at random from a distribution of stimuli, may be categorized as both a correct detection and a false alarm depending on the degree to which the stimulus represents a critical event. For instance, a convenient range for a stimulus dimension is one in which the strength of the stimulus varies from 0 (a definite non-signal) and 1 (definitely a signal), with a signal value of .5 representing maximal uncertainty in the stimulus itself. That is, a stimulus with a signal value of .5 has properties of both a non-signal and a signal. Implicit in this model is the assumption that signal uncertainty exists not only within the observer (a major insight provided by traditional SDT) but in the stimulus dimension itself. This experiment represents an initial test of FSDT using a perceptual signal detection task. In this study fuzzy stimulus and response dimensions were established (seven categories for each), and difficulty (temporal) and bias (instructional set) were manipulated to investigate the degree to which the FSDT model is sensitive to these classical SDT manipulations. The FSDT analysis was compared to an analysis using traditional SDT, using methods outlined in MacMillan and Creelman (1991; see also Macmillan, 2002).

Method

Participants

Seven students from the University of Central Florida (3 men and 4 women, mean age = 21 years) volunteered to participate in this study. Volunteers were compensated by extra credit in their psychology course and were told that a $25 award will be given to the two best performers.

Stimuli

In each trial, the event rate of the stimuli was 21 event/min. Each of the 7 different stimuli was presented 18 times. The order in which the stimuli were presented was randomized.

Procedure

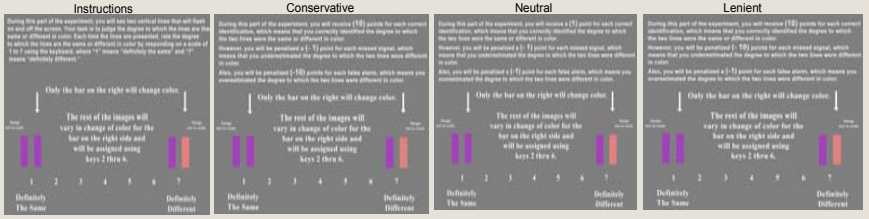
A computerized target detection task, that required color discrimination, was used to collect performance data (d' and c)

After instructions, participants completed a practice trial of the task. Both difficulty levels of the task (less difficult (80% saturation) vs. more difficult (95% saturation) were presented in the practice session, the order in which the participants received the difficulty levels was counterbalanced.

Participants engaged in six 6-min detection tasks consisting of two difficulty levels and three levels of instruction bias (lenient, conservative and neutral). The order in which the participants received the different treatments was counterbalanced using a Latin Square design.

Payoff Matrix for a 2x3 Repeated Measures Design

Difficulty	Bias by Instruction: Conservative	Neutral	Lenient
High	10 Hit -1 Miss -10 FA	1 Hit -1 Miss -1 FA	10 Hit -10 Miss -1 FA
Low	10 Hit -1 Miss -10 FA	1 Hit -1 Miss -1 FA	10 Hit -10 Miss -1 FA



Results

A 2 (difficulty) x 3 (bias instructions) repeated measures analysis of variance was used to test differences among conditions in perceptual sensitivity (d') and response bias (c). Mean sensitivity scores for each difficulty level and bias instruction using traditional and fuzzy computation methods are presented in Figure 1 and Figure 2, respectively. Sensitivity (d') was calculated using the Traditional SDT method of analysis collapsed across stimulus, total d' (see MacMillan & Creelman, 1991, see also Wickens, 2002).

An ANOVA revealed that participants were significantly more sensitive in the less difficult conditions compared to the more difficult conditions, $F(1, 6) = 21.12$, $p = .004$. Participants did not significantly differ in their d' scores across bias conditions, (p>.05).

An ANOVA of sensitivity scores based on the FSDT method of analysis (collapsed across stimuli) revealed that participants were significantly more sensitive in the less difficult conditions compared to the more difficult conditions, $F(1, 6) = 32.23$, $p = .001$ (see Table 1). In addition there was a significant main effect for bias, $F(2, 12) = 7.97$, $p = .006$ (see Figure 2). Data from pairwise comparisons revealed that sensitivity is greater in the conservative condition compared to the neutral condition and lenient.

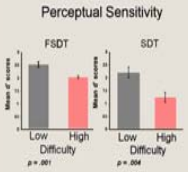


Figure 1. Perceptual sensitivity (d') as a function of difficulty level

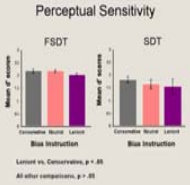


Figure 2. Perceptual sensitivity (d') as a function of bias instruction

When c was calculated using the FSDT method of analysis collapsed across stimuli, the data revealed a significant effect for the response bias manipulation (conservative, neutral, and lenient), $F(2,12) = 8.02$, p=.02. The data displayed in Figure 3, revealed from the pairwise comparisons, that participants' bias (degree of conservativeness and leniency) in purity discrimination was significantly more conservative for stimuli presented in the conservative condition compared to the participants' bias toward stimuli presented in the lenient and neutral conditions, which did not significantly differ from each other.

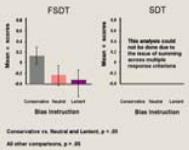


Figure 3. Response bias (c) as a function of bias instruction.

When d' was calculated using the FSDT method of analysis by stimulus, an ANOVA revealed that participants' sensitivity was significantly different across stimuli (2 through 6), $F(3, 16) = 7.61$, $p = .03$, and difficulty levels, $F(1, 6) = 12.604$, $p = .01$. Pairwise comparisons indicated that participants' sensitivity for color discrimination was significantly greater for stimuli with maximal uncertainty such as stimulus 4 ($M = 2.47$, $SD = 0.40$), stimulus 5 ($M = 2.35$, $SD = 0.72$), and stimulus 3 ($M = 1.92$, $SD = 0.24$), and lower for stimuli with minimal uncertainty such as stimulus 2 ($M = 1.91$, $SD = 0.35$) and stimulus 6 ($M = 0.74$, $SD = 0.10$). This trend appears to be stronger for the less difficult condition ($M = 2.05$, $SD = 0.22$) compared to the more difficult condition ($M = 1.70$, $SD = 0.17$).

Similarly, when d' was calculated using the Traditional SDT method of analysis by stimulus interval, the analysis revealed that participants were significantly more sensitive in the less difficult conditions ($M = 2.21$, $SD = 0.20$) compared to the more difficult conditions ($M = 1.23$, $SD = 0.20$), $F(1, 6) = 21.12$, $p = .004$. The data revealed a significant main effect for stimuli (i.e., stimuli 1 to 2, 2 to 3, 3 to 4, 4 to 5, 5 to 6, and 6 to 7), $F(5, 30) = 5.26$, $p = .001$, indicating that participants' sensitivity in color discrimination was significantly different for stimuli intervals. However, in contrast to the findings from the FSDT method of analysis, the traditional analysis indicated that participants discriminated stimulus 6 from 7 ($M = 3.64$, $SD = 0.70$) with the highest sensitivity in color discrimination, followed by the interval between stimuli 2 and 3 ($M = 2.06$, $SD = 0.22$), stimuli 3 and 4 ($M = 1.77$, $SD = 0.34$), stimuli 4 and 5 ($M = 1.64$, $SD = 0.39$), and stimuli 5 and 6 ($M = 1.20$, $SD = 0.38$). Participants discriminated stimuli 1 and 2 with the least sensitivity ($M = 0.03$, $SD = 0.72$).

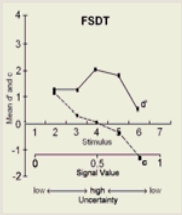


Figure 4. Fuzzy SDT method of analysis by stimulus for d' and c.

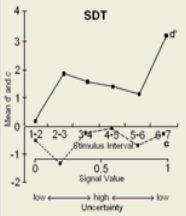


Figure 5. SDT method of analysis by stimulus for d' and c.

Conclusions

Fuzzy Signal Detection analysis and Traditional Signal Detection provide different vistas into signal detection performance. That FSDT provided a better description of the effects of stimulus uncertainty on observers' response bias and sensitivity is not surprising, given that the FSDT model explicitly captures this uncertainty.

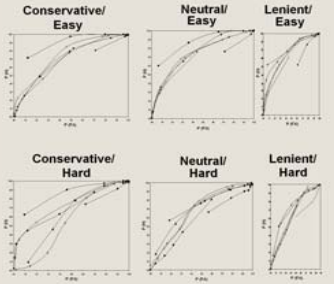
However, the finding that the Traditional SDT model and the FSDT model were perceptually sensitive (d') to different stimuli is interesting. However, it should be recognized that the meaning of d' is different for FSDT model compared to the Traditional SDT model.

The sensitivity index, in the Traditional SDT model, is gauging the participants' ability to place stimuli on the signal dimension, while for FSDT, the index captures sensitivity to the degree of signalness. For FSDT results indicated that participants correctly gauged the degree of signalness and were more sensitive to stimuli in the middle, with the most uncertainty. For Traditional SDT results indicated that participants were most sensitive to the stimulus at the high end of the dimension.

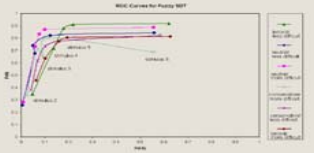
The individual results are not surprising because the FSDT model captures uncertainty in the data and this explains why the data reveals the highest sensitivity for stimuli with the most uncertainty. The results associated with the Traditional SDT analyses conform to the psychometric function of the stimulus dimension. Because the traditional methods are from of multi-dimensional scaling, such a result is not surprising (MacMillan and Creelman, 1991; Wickens, 2002). The surprise is in the very different results that are seen when with the same data set is analyzed using FSDT versus Traditional SDT.

The different findings suggest that further work is necessary to investigate the relation between the index of sensitivity and the FSDT model. These results do confirm, however, that FSDT can be used to analyze system performance in domains in which stimulus categories vary along a continuum. Further, initial evidence suggests that Traditional SDT and FSDT methods may provide different insights into system performance.

ROC Curves for Traditional SDT



ROC Curves for Fuzzy SDT



References

Hancock, P.A., Masalonis, A.J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science*, **1**, 207-230.

Indow, T. & Stevens, S.S. (1966). Scaling of saturation and hue. *Perception & Psychophysics*, **1**, 253-271.

Macmillan, N.A., & Creelman, C.D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Myers, J.L., & Well, A.D. (1995). *Research Design and Statistical Analysis*. (2nd Ed.). New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, **42**, 636-659.

Snodgrass, J.G., & Corwin, D. (1988). Pragmatics of measuring recognition memory. Application to dementia and amnesia. *Journal of Experimental Psychology, General*, **117**, 34-50.

Wickens, T.D. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, P.A. Hancock, principal investigator. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

PERFORMANCE, WORKLOAD, AND STRESS CORRELATES OF TEMPORAL AND SPATIAL TASK DEMANDS

Jennifer M. Ross, James L. Szalma, Jennifer Thropp, and Peter A. Hancock

Department of Psychology and Institute of Simulation and Training
University of Central Florida
Orlando, FL.

Increased understanding of the mechanisms by which stress impacts performance is essential to the design and operation of complex information systems. This study represents a test of the hypothesis of Hancock, Szalma, and Weaver (2002) that the attentional narrowing observed under stressful conditions results from spatial and temporal perception drawing on common resource capacities. Although the present results were unable to resolve the specific issue to a satisfactory degree, a novel finding was observed that noise increases leniency in responding. The impact of noise on performance thus depends on the characteristics of the task to be performed, with spatial uncertainty exerting a significant influence on perceived workload.

INTRODUCTION

The distortion of spatial perception under stress was first formalized by Easterbrook (1959) as resulting from restriction in the range of cue utilization resulting from emotional arousal. This formulation was aimed at explaining the 'inverted-U' hypothesis regarding arousal (including stress) and performance. While the notion of unitary arousal has slowly dissolved, evidence for the narrowing of attention under stress has been experimentally validated (e.g. see Cornsweet, 1969; Dirkin & Hancock, 1984). Recently, Hancock and Weaver (2003) have argued that the distortion in the perception of time results from the same mechanism as narrowing of spatial perception, and that this narrowing is therefore a general effect of stress on perception of space-time.

It was also argued by Hancock, Szalma, and Weaver (2002) that hypothetically this results from perceptions (and distortions of those perceptions) of spatial and temporal information drawing on common resource capacities. These narrowing effects should be therefore exacerbated in the presence of an external stressor such as noise. In addition to performance effects, perceived workload and perceived stress level should increase when a task requires simultaneous processing of spatial and temporal information. It has been well established that experimental manipulations that impact performance in target detection tasks also influence perceived workload and self-reports of stress in those tasks (e.g., Warm, Dember, & Hancock, 1996; Szalma, 2002). Performance decrements are often associated with increases in workload (although see Parasuraman & Hancock, 2001; Yeh & Wickens, 1988), and contexts in which performance declines occur are associated with increases in self-reports of stress (e.g. Matthews et al., 2002; Szalma, 2002). The present experiment represents part of an on-going program to test more hypotheses regarding the relation between changes in the perception of spatial and temporal processing under conditions of high stress and workload. Specifically, a target detection task was employed in which the relative strength of the spatial and temporal components was varied, such that these components were processed either separately or in combination. If spatial and temporal processing draw on similar resource capacities, performance decrements should be observed when the two components are combined as compared to tasks in which response on only one dimension is required.

METHOD

Twenty-two females and twenty-two males, at the University of Central Florida volunteered to participate. They ranged in age from 18 to 32, with a mean of 21.27 years.

Three tasks were employed in the present study. The order of task performance was counterbalanced. In each task participants were required to discriminate a 2mm diameter O from a D and a "backwards-D," which were presented on a background mask of .4 mm diameter white circles on a black background. The tasks differed according to spatial and temporal properties of target detection. The spatial-only task required observers to detect an O of brighter illumination

that could appear in any one of nine locations on a screen. Thus the spatial-only task presented observers with spatial uncertainty regarding target location. The 'temporal-only' required a temporal discrimination instead of an intensity discrimination, and observers were required to report when they observed an O of shorter duration (300 msecs). Non-targets were presentations of an O of longer duration (500 msecs), or a D or backwards D of either duration. In this condition however, there was no spatial uncertainty; stimuli were always presented in the center of the screen. In the third task ('combined') the temporal discrimination was combined with spatial uncertainty. Thus, observers were required to detect an O of shorter duration that could appear in any one of nine locations on the screen. Note that the number of features in each conjunction search was constant across the three tasks.

The presentation rate was 24 events per minute. In the conditions with spatial uncertainty the participant responded by pressing a key on the numerical keypad, from 1-9, corresponding to the quadrant on the screen in which the target appeared. In the temporal-only task, participants responded by pressing the five key on the numerical keypad. Correct responses occurring within 2.5 seconds of the onset of a signal were recorded as Correct Detection; all other responses were recorded as False Alarms.

For the noise group an 85 dBA intermittent white noise was presented via headphones during the three seven minute blocks. The intermittent white noise was composed of ten 4-second, eight 6-second, five 8-second, and four 10-second durations. The order and times of administration of noise was randomized across each session. The total duration of white noise per task was 2.8 minutes. Observers in the quiet condition also wore headphones to control for the effects of additional stimulation. The ambient sound level in the quiet condition was 64.5 dBA.

Three levels of task were factorially combined with two levels of noise to yield six experimental conditions. Participants were randomly assigned to either the noise or quiet condition. Each participant experienced each of the three tasks, and was assigned at random to one of the six possible task orders.

Perceived workload was assessed using the NASA Task Load Index (TLX), self-reports of stress were measured using the Dundee Stress Questionnaire (DSSQ; Matthews et al, 2002). The DSSQ is composed of eleven scales reflecting the observer's cognitive and emotional states. These include Tense Arousal, Hedonic Tone, and Confidence and Control, reflecting the individual's degree of distress; Energetic Arousal, Concentration, and two kinds of motivation (Intrinsic and Success) which indicate the individual's degree of task-engagement; and Self-Esteem, Self-Focused Attention, and two forms of cognitive interference (Task-Irrelevant and Task-Related) which reflect the individual's degree of worry.

Following the instructions for the first task participants completed the pre-DSSQ, after which they began the first task. Upon completion of each task participants were administered the post-DSSQ and the NASA-TLX, which were administered in a randomly assigned counterbalanced order. This order was repeated after the second and third tasks.

RESULTS

The correct detection and false alarm rates for each participant were used to compute signal detection theory measures of sensitivity (d') and response bias (c) in each minute of task performance. Sensitivity scores were analyzed using a 2 (noise) x 3 (task) x 6 (minutes on task) mixed ANOVA with repeated measures on the last two factors. In all test involving repeated measures factors Box's epsilon was used to adjust the degrees of freedom for violations of the sphericity assumption (Myers & Well, 1995). A significant effect for task, $F(2,68)=6.23$, $p<.01$, was observed, and a significant effect for time on task was also evident, $F(4,181)=2.41$, $p=.05$. All other effects failed to reach significance ($p>.05$ in each case). Post-hoc comparisons revealed that observers were significantly more sensitive in the temporal-only task (M=7.495) than in the spatial-only task (M=6.458) and the combined task (M=5.943), but that their scores in the latter two tasks did not significantly differ from one another. Sensitivity scores increased over the first three minutes and declined over the last three minutes across all three tasks (M=6.49, M=7.19, M=6.88, M=6.75, M=6.43, M=6.405, in each minute respectively).

Mean response bias scores for each noise condition are plotted as a function of time on task in Figure 1. ANOVA of the c scores revealed that observers in the noise condition (M=-.93) were significantly more lenient in responding than their cohorts in the quiet condition (M=-.43), $F(1,42)=4.98$, $p<.05$. A marginal effect for time was also observed, $F(4,173)=2.13$, $p=.08$, with a trend for

conservatism to rise over time. All other effects failed to reach significance (p>.05 in each case).

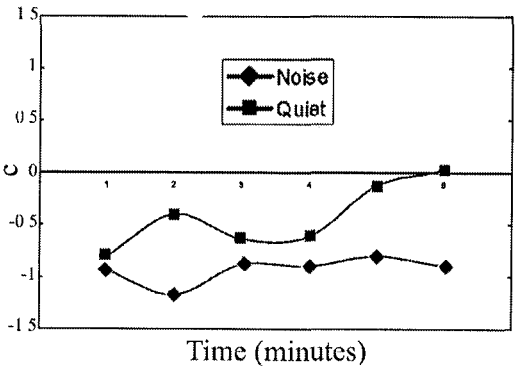


Figure 1. Response bias as a function of time on task for each noise condition.

We rejected any artifact based upon order effects since initial analysis showed that scores on the NASA-TLX did not significantly vary as a function of the order of task performance. The temporal-only task induced the least global workload, and noise induced higher workload in each task. An ANOVA confirmed these impressions, with a significant main effect for task, $F(2, 82)=3.53$, $p<.05$, although the effect for noise was only marginally significant, $F(1, 42)=3.57$, $p=.07$. *Post-hoc* comparisons indicated that the temporal task (M=49.34) induced significantly less global workload than the spatial-only task (M=55.41) and the combined task (M=55.37), which did not significantly differ from one another. In addition to global workload, weighted ratings on the subscales of the TLX were obtained. Due to the paired-comparison procedure used to obtain the weights, physical demand was excluded from the analysis to meet the independence assumption of the statistical test. Physical Demand was selected because it contributed least to overall workload (see Warm, Dember, and Hancock, 1996). A 2 (noise) by 3 (task) by 5 (subscale) ANOVA revealed a significant effect for noise, $F(1, 42)=4.30$, $p<.05$, with the noise condition inducing significantly higher workload (M=167.71) than the quiet condition (M=140.97). Significant effects were also observed for scales, $F(3, 124)=9.72$, $p<.0001$, and a task by scales interaction, $F(6,242)=2.45$, $p<.05$. Separate task by noise ANOVA'S for each scale were computed. Significant effects for task were observed for the Weighted Temporal Demand subscale, $F(2, 83)=5.90$, $p<.01$, and post-hoc comparisons indicated that the combined task induced more temporal demand (M=169 30) than either the spatial-only (M=118.75) or temporal-only task (M=133.75), which did not significantly differ from one another. In addition, a marginal noise by task interaction was obtained, $F(2, 83)=2.67$, $p=.08$, with a trend for noise to induce more temporal demand in the spatial-only task A significant effect for task was also observed in the Weighted Performance subscale, $F(2, 78)=3.61$, $p < .05$, and post-hoc tests revealed that the combined task (M=89.43) induced significantly less Performance Workload than the spatial-only task (M=127..57). A marginal effect for noise was observed in the Weighted Frustration subscale, $F(1, 42)=3.91$, $p=.06$. with noise having higher frustration levels (M=166.82) than quiet conditions (M=91.30).

Participant's responses on the pre- and post-DSSQ were used to calculate scores on eleven scales using the formula (post-score – pre-score/SD), where the standard deviation use was obtained from a large normative group obtained from Matthews et al. (1999), except the two motivation scales, for which normative values are not available. For these scales, sample standard deviations were used (Matthews, Campbell, & Falconer, 2001). Note that preliminary tests revealed that the order of task performance did not exert a significant influence on DSSQ change scores. Separate 2 (Noise) by 3 (Task) ANOVAs were computed for each scale. Significant task effects were observed for Energetic Arousal, $F(2, 75)=5.82$, $p<.01$, and post-hoc tests indicated that the temporal-only task (M=-0.44) induced a significantly larger drop in Energetic Arousal relative to pre-task state as compared to the spatial only (M=-0.16) and combined (M=-0.15) tasks, which did not significantly differ from one another. Tense Arousal had a significant main effect for noise, $F(1, 42)=3.98$, $p=.05$. Observers in the noise condition reported significantly higher Tense Arousal (M=0.74) than observers in the quiet condition (M=0.20). A significant noise effect was also observed for the Hedonic Tone scale, $F(1, 42)=6.84$, $p=.01$. Participants in the noise condition showed a large decline in Hedonic Tone (M=-0.90) than observers in the quiet condition (M=-0.20). Further, a significant main effect for task was observed, $F(2, 83)=3.07$, $p=.05$. Post-hoc tests revealed that hedonic tone was significantly lower after a temporal task (M=-0.67) than in the combined task (M=-0.46), but that change scores after the spatial-only task (M=-0.52) did not differ significantly from those associated with the other two tasks. For Intrinsic

Motivation there was a marginally significant main effect for task, $F(2, 72)=2.96$, $p=.07$, with temporal-only task ($M=-0.67$) tending to show greater declines in Intrinsic Motivation than those in the combined task ($M=-0.38$). However, there was also a marginal Noise by Task interaction, $F(2, 72)=3.04$, $p=.07$, which showed a tendency for noise to reduce Intrinsic Motivation, but only in the spatial-only task.

DISCUSSION

Due to a potential ceiling effect in the detection rate itself, the performance data did not provide unequivocal information either for or against the hypothesis of Hancock et al. (2002) regarding the covariation of temporal and spatial distortion. Thus the only difference in sensitivity was observed in the expected superiority of a task without spatial uncertainty. These effects extended to perceived workload, which was higher in tasks with spatial uncertainty, regardless of whether one of the discriminations was temporal or non-temporal. Similarly, both tasks with spatial uncertainty reduced Energetic Arousal and Intrinsic Motivation (in the case of the quiet condition) relative to the temporal task without such uncertainty. As noted, the tasks employed here were very easy, with average d' scores above 5 for all three tasks. It may be that differences in performance would be observed with more difficult discriminations, and efforts are currently underway to investigate this possibility

However, two subscales of the TLX and three subscales of the DSSQ indicated that task type had effects on subjective ratings that differed from those described above. Thus, results with the Weighted Temporal Demand scale showed that while spatial uncertainty (spatial-only task) and temporal discrimination without uncertainty (temporal-only task) induced similar Weighted Temporal Demand, combining these features (combined task) increased Weighted Temporal Demand, although a marginal interaction effect between task and noise suggested that this effect may have been stronger in the quiet condition than in the noise condition. In contrast, participants in the combined condition rated their performance as better in comparison to the spatial-only task, indicating a facilitative effect of adding a temporal discrimination to a task with spatial uncertainty. The latter effect accords with the main effect for task observed in the Hedonic Tone scale of the DSSQ, in which observers reported lower hedonic tone after the temporal-only task than after the combined task, regardless of the order in which the tasks were performed. Similarly, a marginal effect for intrinsic motivation indicated that the temporal-only task induced greater reductions in intrinsic motivation relative to the combined task. Contrary to expectations, combining the temporal and spatial elements of a task reduced the effects of task stress along these dimensions. Current experimental work will investigate whether these effects extend to a task in which spatial and temporal discriminations are combined rather than temporal discrimination and spatial uncertainty.

In this study, noise did not impact sensitivity, but it did decrease response bias, indicating a more lenient response bias in the presence of noise. A marginal noise by time interaction suggested a rise in conservatism in the quiet condition and no change over time in the noise condition. These results are consistent with the general view that stress lowers response criteria (Welford, 1973), but not with the literature on the impact of noise, which indicates that the criterion should increase (e.g. see Broadbent, 1971) or not significantly change (Milosevic, 1983). The consensus, based upon the work of Broadbent (1971; 1978; see also Poulton, 1978), is that in long duration tasks, continuous noise induces a rise in conservatism, particularly for responses of intermediate confidence and that in shorter duration tasks noise may serve to enhance performance. However, in this study, which employed a short duration task, no performance enhancement (in terms of d') was observed, and response bias decreased, indicating increased responding in the presence of intermittent noise. Note that in a short duration vigil using similar stimuli to those employed in this study, (Helton, Warm, Matthews, Corcoran, & Dember, 2002) found that continuous jet engine noise improved detections without changing the false alarm rate. A crucial difference between the study of Helton and his colleagues and the current study was the format of noise. In their study, noise was continuous presentation of jet engine noise, while in the current study intermittent noise of variable duration and random presentation times was used. Thus, participants experienced a greater degree of uncertainty regarding when and for how long noise would be present. It may be that it is uncertainty as a source of stress that induces leniency in responding.

The impact of noise on observers' state was more pronounced in workload and stress, with

participants experiencing noise reporting increases in Weighted Frustration and a marginal increase in Global Workload. Similarly, participants in the noise condition reported more Tense Arousal and lower Hedonic Tone, and in the case of conditions with spatial uncertainty, less Intrinsic Motivation. These effects differ from those of Helton and his colleagues, who observed that noise decreased the symptoms of stress. In their study, noise decreased the anger/frustration experienced by observes after a difficult vigilance task, and increased energetic arousal, intrinsic motivation and confidence and control relative to a quiet condition. The findings in regard to the latter three scales indicate an increase in task engagement in the presence of noise. These results represent trends opposite to those observed in this study, despite the fact that the task demands placed upon operators by Helton et al. were substantially greater than those in this study. For instance, the event rate in their study was 57.5 events per min, while an event rate of 24 events per minute was employed herein. It may be that whether noise exerts a negative effect on stress and performance depends on how the difficulty level of the task interacts with the noise presented to operators.

ACKNOWLEDGEMENT

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. P.A. Hancock, *Principle Investigator.* The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The assistance of the Psychology Department at University of Central Florida is also gratefully acknowledged.

REFERENCES

Broadbent, D.E., (1971). *Decisions and stress*. England: Medical Research Council.

Broadbent, D.E., (1978). The current state of noise research: Reply to Poulton. *Psychological Bulletin,* **85**(5), 1052-1067.

Cornsweet, D.M., (1969). Use of cues in the visual periphery under conditions of arousal. *Journal of Experimental Psychology,* **80**(1), 14-18.

Dirkin, G.R., & Hancock, P.A. (1984). Attentional narrowing to the visual periphery under temporal and acoustic stress. *Aviation, Space and Environmental Medicine,* **55**, 457.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review,* **66**, 183-201.

Hancock, P.A., Szalma, J.L, & Weaver, J.L. (2002). Distortion of perceptual space-time under stress. Paper presented at the 23rd Army Science Conference, Orlando, FL, December.

Hancock, P.A., & Weaver, J.L. (2003). On time distortion under stress. *Theoretical Issues in Ergonomic Science,* in press.

Helton, W.S., Warm, J.S., Matthews, G., Corcoran, K.J., & Dember, W.N. (2002). Further tests of an abbreviated vigilance task: Effects of signal salience and jet aircraft noise on performance and stress. Proceedings of the Human Factors and Ergonomics Society, **46**, 1546-1550.

Matthews, G., Campell, S.E., & Falconer, S., (2001). Assessment of motivational states in performance environments. Proceedings of the Human Factors and Ergonomics Society, **45**, 906-910.

Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., et al. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. Emotion, **2**(4), 315-340.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards the state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds), *Personality psychology in Europe* (vol 7), (pp. 335-350).

Milosevic, S., (1983). Effects of noise on signal detection. *Ergonomics*, **26**(10), 939-946.

Myers, J.L., & Well, A.D., (1995). *Research design and statistical analysis*. Hillsdale, NJ: Erlbaum.

Parasuraman, R., & Hancock, P.A., (2001). Adaptive control of mental workload. In: P.A. Hancock and P.A. Desmond (Eds.) *Stress, workload, and fatigue.* (pp 305-320), Mahwah, NJ: Erlbaum.

Poulton, E.C., (1978). A new look at the effects of noise: A rejoinder. *Psychological Bulletin, 85(5), 1068-1079.*

Szalma, J.L. (2002). Workload and stress of configural displays in vigilance tasks. Proceedings of the Human Factors and Ergonomics Society, **46**, 1002-1006.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In: R. Parasuraman and M. Mouloua. (Eds). *Automation and human performance*: Theory and applications. (pp 183-200), Erlbaum, Hillsdale, N.J.

Welford, A.T., (1973). Stress and performance. *Ergonomics*, **16**, 567-580.

Yeh, Y., & Wickens, C.D., (1988). Dissociation of performance and subjective measures of workload. *Human Factors,* **30**, 111-120.

THE STRESS-PERFORMANCE RELATIONSHIP AND THE EFFECT OF PERFORMANCE MODERATORS

Paul Ward, James L. Szalma, and Peter A. Hancock

Department of Psychology and Institute for Simulation and Training, University of Central Florida

Hancock and Warm (1989) outlined a dynamic model of stress and performance detailing an individual's psychological and physiological adaptivity to stressful environments. The model predicts that psychological adaptivity (e.g., attentional narrowing) is exhausted in a similar manner to the way in which stress degrades physiological response capacity. Although there is usually more than one form of input stress, task demand is proposed to be the primary source of stress. Recent predictions indicate that stress may affect tasks differently, depending upon the nature of the task. Without a comprehensive evaluation of the literature, prediction of how task performance will be debilitated or facilitated by stress, or whether stress effects can be alleviated remains to be fully answered. Current modeling efforts have attempted to predict soldier-system performance. However, these are not without limitation. Our proposed descriptive and meta-analytic framework attempts to specifically address some of these limitations.

Introduction

In preceding years, researchers have examined stress primarily using one of three inter-related approaches. Whilst some have considered stress as a property of the environment (e.g., Hockey, 1983), others have examined the appraisal or coping mechanisms of stress (e.g., Lazarus, 1966), or the associated physiological response (Selye, 1956). Researchers have often embarked upon these approaches independently of each other. In addition, the majority of research has concentrated upon the effects of a single stressor on 'isolated' information processes. Researchers have typically not examined the variable and/or predictive effects of multiple stressors on a range information processes. In one exception, Hockey and Hamilton (1983) examined the cognitive patterning of stress states (e.g., noise, narcotic, etc.) on strategic (i.e., selective attention, decision-making) and structural (i.e., short-term memory, alertness) processes. This research highlighted that while some commonality exists between particular stressors, a multidimensional approach was necessary to explain the lack of response uniformity.

Lazarus and Folkman's (1984) focus on mechanisms of appraisal and coping has emphasized the person-environment-stress interaction. This viewpoint suggests that the transaction between an individual and their appraisal of the environmental task as taxing or exceeding available resources moderates the effect of a stressor. Consistent with Hancock and Warm (1989), Matthews (2001) indicated that this transaction occurs at multiple levels (e.g., biological, psychological).

In reality, however, the environmental, coping and physiological response approaches are three aspects of a single dynamic process which could be used to describe responses at multiple levels of analysis (Hancock & Warm, 1989). More specifically, these authors suggested that the study of the physical characteristics of environmental stressors was a deterministic, input approach to examining the influence of stress upon performance, whereas coping mechanisms reflected a nomothetic, compensatory or adaptive approach, and bodily response, an ideographic, output approach.

The Model

The Hancock and Warm (1989) model, first presented as a dynamic model of stress and sustained attention, outlined three modes of operation reflective of adaptational success, from dynamic stability to dynamic instability. Hancock and Warm predicted that at minor levels of input stress, behavior is not adversely affected, and thus, no output stress would be observed. These low levels of stress are adequately dealt with by an individual's adaptive capability. On a continuum

ranging from hypostress to hyperstress, the central point, the *zone of comfort*, reflects such adaptive capability to minor stress levels. However, as the intensity of, or exposure to, stress rises (e.g., towards the region of *maximal adaptability*), and adaptive capability is eventually surpassed, output is affected. The result is increasing discomfort, a reduction in performance efficiency and then a rapid decrease in adaptive capability. Importantly, the model indicates that input stressors affect both psychological (i.e., resources allocation) and physiological adaptability (i.e., homeostatic regulation). Whilst, an input stressor may appear to affect psychological processes alone, physiological functioning is also typically indirectly affected. Moreover, the region of maximal adaptability is larger for physiological functioning. Although attentional resources are depleted when maximal psychological adaptability is exceeded, a swift decline in physiological adaptability closely follows.

The model is defined by two different constituents of information, *information rate*; the temporal flow of environmental information, and *information structure*; the perceived meaning of the information's content. The latter is specifically influenced by the task, stressor and future expectations. Although there is usually more than one form of input stress the primary source is task demand (Hockey, 1986; Hancock & Warm, 1989). Moreover, Hockey and Hamilton's research illustrates that variable effects of stressors are likely to be observed as the nature of the information processing requirements of the task change.

In a normal information-rich environment and under low levels of stress, multiple solutions to a task may be available, and different individuals are likely to interpret information in a diverse number of ways in achieving a successful outcome. The result is an array of functionally operational goal-directed behaviors. However, as stress increases, the strategies available to an operator that lessen task demand are reduced, and the number of solution paths with respect to the goal is diminished. Consequently, the stereotypical nature of behavior is increased as attention narrows to finding the most appropriate and available solution to the task. During task performance, individuals seek to maximize their capacity to retain optimum information throughput using such compensatory strategies as 'attentional narrowing' to modify the perceived rate and structure of information.

Theoretical Predictions and Outstanding Issues

A number of theoretical predictions can be made both from the model and recent works elaborating on this theme (Hancock, Szalma & Weaver, 2003; Szalma & Hancock, 2003; Ward, Szalma & Hancock, 2003). Of particular interest here, however, is the way in which performance degrades over time under stressful conditions and how performance is restored to within a normative range. The model predicts that the drain on attentional resources (e.g., narrowing) occurs in a similar manner to the way in which stress degrades physiological response capacity. However, recent predictions derived from the model indicate that stress may affect tasks differently, depending upon the nature of the task and/or how well the task is learnt. Performance is predicted to degrade quicker for those tasks that are more knowledge-driven than those that are more skill-oriented (cf. Rasmussen, 1983). Moreover, the degree to which a task is automated or proceduralized (see Anderson, 1983; Fitts & Posner, 1967); a product of learning and skill development, may determine the relative effect of stress on performance. Lastly, it is unknown whether there is a hysteresis effect with respect to depletion and recovery of attentional resources dependent upon the knowledge-, rule-, or skill-oriented basis of the task. An implication is that if there are differential effects of stress depending upon the nature or constituents of the task, then current computational tools used to model performance effects may need to take these factors into consideration during, rather than after the modeling process.

Detailing the multidimensional nature of stress and performance is an onerous task. However, without a comprehensive evaluation of the literature, prediction of how performance will be debilitated or even facilitated by stress will be even more arduous. What's more, this task is considerably amplified when one considers the vast range of potential interactions between different environmental, task-based and social stressors and the information processing requirements of tasks sub-serving perception, cognition and action. A

recent review by Pew and Mavor (1998) suggested that only limited information is available on the effects of all known environmental stressors. These authors summarized known effects for heat, toxic substances, noise, vibration, as well as factors related to fatigue (e.g., sleep degradation), and cognitive workload. While the details of the effects are beyond the scope of this paper, Pew and Mavor concluded that much of the current research has been conducted with the aim of eliciting threshold limits as opposed to determining the degree of degradation per unit of time.

When one includes those factors that may moderate or mitigate the stress-performance relationship, the complexity of performance prediction becomes all too evident. For instance, genetic predisposition may make some individuals more susceptible to stress, or at least to particular stressors, than others. Those who are high in trait anxiety typically demonstrate a greater propensity toward the detecting threatening stimuli and interpreting neutral stimuli as threatening. These individuals are also more likely to allocate a greater number of resources to worry and task-irrelevant processing increasing the probability of performance impairment (Eysenck 1992, Williams, Watts, McLeod, & Matthews, 1988). The transaction between a stressor and an individual's appraisal of the event is likely to play a significant role in output levels of stress and both efficiency and effectiveness of task performance. However, depending upon the nature and intensity of the stressor, even compensatory or adaptive processes are likely to have little effect once an individual reaches the region of *maximal adaptability*. In comparison, personal characteristics such as other emotions, intelligence or cognitive style, or alternatively, the amount of training received or level of skill acquired may alleviate the known stress effects (for a summary, see Pew and Mavor, 1998). Furthermore, learned strategies which allow an individual to maintain performance in the comfort zone for longer periods or at least, delay the onset of decline beyond the region of maximal adaptability are likely to facilitate performance in the presence of a stressor that usually has debilitating effects.

The Applied Problem

From a practical perspective, one often simply wants to know how stress affects performance and by how much performance will typically be degraded. Moreover, the question often arises, is there anything one can do to alleviate these effects? Of related interest are whether particular individuals are more suited to the task than others? Can we aid in the selection of best man for the job? Can we train the right person to acquire the appropriate strategies? In order to answer each of these questions the stressor-task-moderator relationship outlined above needs to be fully delineated.

IMPRINT (Improved Performance Research Integration Tool), a task network modeling architecture used to predict soldier-system performance under a myriad of conditions, is one such system which has attempted to begin to answer some of these questions. One of the major uses of IMPRINT is to determine the soldier performance requirements for system procurement purposes, and essentially to explore 'what if' scenarios (Allender, Salvi, & Promisel, 1997; Lebiere, Biefeld, Archer, Archer, Allender, *et al.*, 2002; O'Brien, Simon, & Swaminathan, 1992). Performance shaping functions (PSFs) are applied to modeled data to predict relative change in performance for a specific set of personnel characteristics and training regimen, whereas stress degradation algorithms (SDAs) are used to modify baseline measures to account for negative affects upon performance. Application of each algorithm is determined by task classification (e.g., taxons) assignment so that the same algorithm is applied to weighted tasks that fall within the same taxon. Nine taxons have been identified spanning perceptual, cognitive, psychomotor and communication modalities (see Fleishman, & Quaintance, 1984). Although parsimonious, there is a high likelihood that the current taxonomy conceptualizes the majority of processes as essentially equivalent (e.g., all information processing and problem solving tasks are treated similarly) and consequently, the model may have little explanatory power.

At present, five, predominantly environmental, stressor algorithms exist within IMPRINT. These relate to Mission Oriented Protective Posture gear for biological, chemical and nuclear defense, heat, cold, noise, and sustained operation (Allender,

Salvi, & Promisel, 1997). However, much of the research used to derive these algorithms is dated and the current depiction does not capture the state of the art with regards to known stressor effects. Moreover, information for each of the five stressors is not available for all taxons and a comprehensive evaluation of current stressors has not been implemented.

To increase the capability of modeling and predicting human cognition, steps are currently being taken to integrate IMPRINT with ACT-R (Adaptive Control of Thought – Rational; Anderson & Lebiere, 1998), a hybrid cognitive architecture. ACT-R can generate distributions of time and error data for each task within IMPRINT that appropriately reflect the actual perceptual, cognitive, and motor components of tasks experienced in the real world, albeit under standardized conditions (for detailed summary, see Lebiere et al., 2002). However, current efforts do not appear to appropriately capture the transactional viewpoint by excluding stress degradation algorithms from the cognitive modeling processes. Rather, the interactive and adaptive stress process (e.g., SDAs) as well as the influence of moderating variables (e.g., PSFs) are applied post hoc – clearly at a point when the process of affect or mitigation is rendered largely superfluous, or at best, provides only a poor estimate of degradation. To the best of our knowledge, there is no plan to model the stress process in ACT-R.

The Solution

The limited number of SDAs currently incorporated within IMPRINT indicates that the effects of other environmental, social, and task-based input stressors need to be examined. More specifically, how these stressors affect a wide range of information processes sub-serving perception, cognition and action is of particular concern. Examination of these processes will aid in detailing the cognitive patterning of stress states in a similar, albeit more complete manner, to that first proposed by Hockey and Hamilton (1983). The interplay between stressors and compensatory mechanisms captured, for example, by the moderator variables described above, also need to be incorporated into a comprehensive system. Such a system will allow the trinity of stress (e.g., input, adaptive, and output processes) to be fully examined and the associated effects to be evaluated. To achieve this aim, we propose a descriptive and meta-analytic framework to collectively address each of these concerns. Our purpose is to detail the psychological and physiological adaptivity with respect to different information processing tasks, stressors, and moderators. In addition to gaining greater understanding of the stress-performance relationship, part of the impetus for this research is to increase current prediction capacity under stressful conditions by informing the construction of PSFs and SDAs for existing and future stressors within IMPRINT. In addition, our aim is to ensure that existing research is compatible with the adopted theoretical framework and modeled data derived via ACT-R. Collectively, this framework provides an integrative database that will drive theoretical development, facilitate future modeling efforts and allow existing models of stress and performance (e.g., Hancock & Warm, 1989) and their associated predictions to be tested.

References

Anderson, J.R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Allender, L., Salvi, L., & Promisel, D. (1997). Evaluation of human performance under diverse conditions via modeling technology. In *Proceedings of Workshop on Emerging Technologies in Human Engineering Testing and Evaluation*, June 1997. Brussels, Belgium: NATO Research Study Group.

Anderson, J.R. & Lebiere (1988). *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum.

Eysenck, M.W. (1992). Anxiety: The cognitive perspective. London: Lawrence Erlbaum Associates.

Fleishman, E.A. & Quaintance, M.K. (1984). *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic Press.

Fitts, P.M. & Posner, M.I. (1967). Human Performance. Westport, CT: Greenwood Press.

Hancock, P.A. & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.

Hancock, P.A., Szalma, J.L., & Weaver, J.L. (2003). The distortion of perceptual space-time under stress. *Unpublished white paper.*

Hockey, G.R.J. (1983). Stress and fatigue in human performance. New York: Wiley.

Hockey, G.R.J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue and circadian rhythms. In K.R. Boff., L. Kaufman, and J.P. Thomas (Eds.), *Handbook of perception and human performance* (44, pp.1-49). New York: Wiley.

Hockey, R. & Hamilton, P. (1983). The cognitive patterning of stress states. In G.R.J. Hockey (ed.), *Stress and fatigue in human performance* (pp.331-362). New York: John Wiley & Sons Ltd.

Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York: McGraw-Hill.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer Verlag.

Lebiere, C., Biefield, E., Archer, R., Archer, S., Allender, L., *et al*. (2002). IMPRINT/ACT-R: Integration of a task network modeling architecture with a cognitive architecture and its application to human error modeling.

Matthews, G. (2001). Levels of transaction: A cognitive science framework for operator stress. In P.A. Hancock, & P.A. Desmond (Eds.), *Stress, workload, and fatigue.* Mahwah, NJ: Erlbaum.

O'Brien, L.H., Simon, R., & Swaminathan, H. (1992). Development of the personnel-based system evaluation aid (PER-SEVAL) performance shaping functions. *ARI Research Note 92-50*, United States Army Research Institute for the Behavioral and Social Sciences.

Pew, R.W., & Mavor, A.S. (1998). *Modeling human and organizational behavior: Applications to military simulations.* Washington DC: National Academy Press.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics,* vol.SMC-13, 257-266.

Szalma, J.L., & Hancock, P.A.(2003). On mental resources and performance under stress. *Unpublished white paper.*

Selye, H.A. (1956). *The stress of life.* New York: McGraw-Hill.

Ward, P., Szalma, J.L., & Hancock, P.A. (2003). Stress, human information processing, and performance mediators: Application of a descriptive framework to current modeling tools. *Unpublished white paper.*

Williams, J.M.G., Watts, F.N., McLeod, C. & Matthews, A. (1988). *Cognitive psychology and emotional disorders*. Chichester: Wiley.

The Stress-Performance Relationship and Performance Mediators: Evidence from a Meta-Analysis of Noise and Memory

Paul Ward, James L. Szalma, Peter A. Hancock, Peter I. Terrence, April Rose Panganiban, and Justin Fox Morgan

Department of Psychology and Institute for Simulation and Training, University of Central Florida

Introduction

Detailing the multidimensional nature of stress and performance, and predicting how performance is debilitated or facilitated by stress can be arduous task. This task becomes more and more difficult when considering the potential interactions between environmental, task-based and social stressors and the information processing requirements of tasks sub-serving perception, cognition and action. The complexity of performance prediction increases exponentially when one includes factors that moderate the stress-performance relationship, and permit performance effectiveness to be maintained, albeit at the potential cost of processing efficiency.

The majority of research to date has primarily focused upon the effects of a single environmental stressor on 'isolated' information processes (Hockey & Hamilton, 1983). Moreover, the effects of task-based stress have seldom been considered (Hancock & Warm, 1989). While limited information is available on the effects of environmental stress such as heat, toxic substances, noise, and vibration much of the research has attempted to elicit threshold limits as opposed to determining the degree of degradation per unit of time making performance prediction difficult (for a review, see Pew & Mavor, 1998).

IMPRINT

From a practical perspective, the questions often posed are simply: how does stress affect performance, by how much does performance typically degrade, and is there anything one can do to alleviate these effects? IMPRINT (Improved Performance Research Integration Tool), a task network modeling architecture used to predict soldier-system performance under a myriad of conditions, is one system which has attempted to begin to answer such questions in a military context (see Allender et al., 1997, Lebiere et al., 2002). Within IMPRINT, performance shaping functions are applied to modeled data to predict relative change in performance for a specific set of personnel characteristics and training regimen, whereas stress degradation algorithms are used to modify baseline measures to account for negative affects upon performance.

At present, five, predominantly environmental, stressor algorithms exist within IMPRINT (see Figure 1). However, much of the research used to derive these algorithms is dated and the current depiction does not capture the state of the art with regards to known stressor effects. We propose a descriptive and meta-analytic framework to collectively address this shortfall. 'The matrix' is a three-dimensional database which has been constructed to inform the functional content, and where necessary, modify the structural basis of current modeling applications used in soldier-system performance prediction. The impetus for this research is to increase current prediction capacity under stressful conditions by informing the construction of stress degradation algorithms for existing and future stressors within IMPRINT. In the first instance, our aim was to quantify the known effects of an existing stressor within IMPRINT to demonstrate the utility of the system.

Acknowledgement

This research was supported by a grant from the Army Research Organization (Grant# DAAD19-01-1-0621; Principal Investigator, P.A. Hancock). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

A Summit for People & Technology — Human Factors and Ergonomics Society 47th Annual Meeting

Method

Qualitative Description of Stress and Performance Research

The characteristics of experiments examining the stress-performance relationship were descriptively coded and entered into the matrix. Details included sample characteristics, task description, dependent measures and experimental design. To date, over 120 independent experiments from 100 articles have been entered into the matrix across a range of stress and performance indices and mediators of that relationship.

Quantifying Effects

To quantify a sub-set of the studies coded within the matrix we conducted a fixed effects meta-analysis on experiments related to noise stress and memory recall. In total, 28 studies were initially considered for inclusion within the meta-analysis. However, 15 studies had to be excluded due to insufficient data or poor experimental quality. The absence of any pre-post test correlations or individual data precluded calculation of *rho* and so only between group contrasts could be considered. Multiple effects from 16 experiments across 13 studies were included within the final analysis. Based upon an *a priori* grouping scheme effects were categorized on the basis of noise intensity contrasts ranging from >95dBA to <55dBA. The noise group intensities are presented in Table 1. Too few studies were included within the current meta-analysis to examine effects such as noise type (e.g., continuous v. intermittent noise). Future work will aim to specifically examine these issues.

Consistent with the meta-analytic literature, Hedges' g_i (multiplied by $J(m)$) was used as an unbiased estimator of effect size, g_i^u (1) (see Hedges & Olkin, 1985).

Formula: Unbiased estimator of effect size (g_i^u)

$$g_i^u = g_i \times J(m) = g_i \times 1 - \frac{3}{4m-1} \qquad (1)$$

where:

$$g_i = \frac{M_1 - M_2}{s_{pooled}} \qquad s_{pooled} = \sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}} \qquad m = n_1 + n_2 - 2$$

Each effect was then weighted ($g_i^u w_i$) by the reciprocal of the variance to add a greater contribution to the grand mean for those studies with a larger sample size and smaller variance. Mean effect sizes ($T_{i\bullet}$) (2) for each noise intensity group and a grand weighted mean ($T_{\bullet\bullet}$) (3) were calculated (Hedges, 1994; Cooper & Hedges, 1994).

Formulae: Group ($T_{i\bullet}$) and grand ($T_{\bullet\bullet}$) weighted mean effect sizes

$$\overline{T}_{i\bullet} = \frac{\sum g_{ij}^u w_{ij}}{\sum w_{ij}} \qquad (2) \qquad \overline{T}_{\bullet\bullet} = \frac{\sum \overline{T}_{i\bullet} w_{i\bullet}}{\sum w_{i\bullet}} \qquad (3)$$

where:

$$w_i = \frac{1}{v_i} \qquad v_i = \frac{n_1+n_2}{n_1 n_2} + \frac{(g_i^*)^2}{2(n_1+n_2-2)} \qquad w_{i\bullet} = \sum w_{ij}$$

Heterogeneity Test and Removal of Outliers

The heterogeneity statistic, Q, was used to determine whether experiments could be described as sharing a common effect size. Q is the product of Q_W and Q_{BET}, where Q_W provides a test of within-group variability and Q_{BET}, the weighted sum of the squares, is analogous to assessing treatment variance using one-way ANOVA (Hedges, 1994). Heterogeneous groups are likely to necessitate a removal or regrouping of effects (Hedges & Olkin, 1985). In order to reduce heterogeneity, Q_i was recurrently calculated for each effect within the heterogeneous group. Q_i is essentially equivalent to Q_W when the ith effect estimate has been removed from the group.

Results

After the removal of outliers, calculation of the grand mean effect and its 95% confidence interval ($CI_{(1-\alpha)}$) revealed a significant small to moderate negative effect of noise on memory recall, $T_{\bullet\bullet}$ = -0.28, $CI_{(1-\alpha)}$ ± 0.07 (-0.35 ≤ $T_{\bullet\bullet}$ = θ ≤ -0.21). Group 1 (moderate-low *vs.* low) was the only heterogeneous group with each estimate contributing equally to the within group variability. The heterogeneity statistics are summarized in Table 1. The grand mean effect was also recalculated with group 1 removed from the analysis. A significant but small negative effect of noise on memory recall was still found, $T_{\bullet\bullet}$ = -0.17, $CI_{(1-\alpha)}$ ± 0.08 (-0.25 ≤ $T_{\bullet\bullet}$ = θ ≤ -0.10).

Table 1: Heterogeneity Summary Table

Source	Statistic	*df*
Between groups (Q_{BET})	140.58*	47
Between groups (Q_{BET} after heterogeneous group$_i$ was removed)	*89.16**	*23*
Within groups		
Group 1: Moderate-low [65-55dB] *vs.* low [<55dB]	177.98†	23
Group 2: Moderate [78-66dB] *vs.* low [<55dB]	4.26	4
Group 3: High [95-85dB] *vs.* low [<55dB]	3.06	7
Group 4: High [95-85dB] *vs.* moderate-low [65-55dB]	2.54	1
Group 5: High [95-85dB] *vs.* moderate [78-66dB]	3.11	5
Group 6: Very high [>95dB] *vs.* high [95-85dB]	5.67	2
Total within groups (Q_W)	196.63†	42
Total within groups (Q_W after heterogeneous group$_i$ was removed)	*18.65*	*18*
Overall (Q)	337.21*†	5
Overall (Q after heterogeneous group$_i$ was removed)	*107.82**	*4*

***Note:** * denotes p<. 05, heterogeneity is denoted by †*

The group mean effect size estimates (T_i ± $CI_{(1-\alpha)}$) are presented in Figure 2. Q_{BET} indicated significant between group differences when group 1 was included and removed from the analysis. Of the homogeneous groups, group 3 (high *vs.* low), group 5 (high *vs.* moderate), and group 6 (very high *vs.* high) were significantly different from zero.

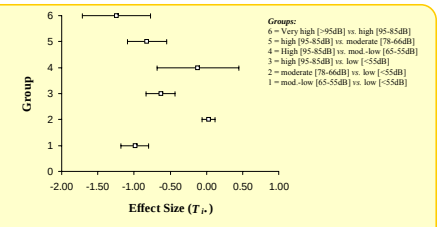


Figure 2: Group Mean Effect Size Estimates (T_i ± $CI_{(1-\alpha)}$).

Bonferroni adjusted post hoc contrasts indicated that there were no differences in the magnitude of the effect between groups 3 to 6, χ^2 (1) = 1.2 to 5.5, p > .01. That is, when contrasting each significant effect, the relative shift in noise intensity (e.g., from low to high, moderate to high, high to very high) is likely to result in a comparably large negative effect on memory recall. Significant differences in the magnitude of effect were found only when contrasting moderate to low noise intensity effects with other effects of a high noise intensity (>85dB) (e.g., high vs. low, χ^2 (1) = 34.64, p > .01). Due to the heterogeneity in the data comparisons with group 1 were not made.

Discussion

While stress can both facilitate and debilitate performance, the negative effect upon memory recall demonstrated by the group and grand weighted mean effect sizes is consistent with the expected direction of effect (i.e., Hockey & Hamilton, 1983). Previously, noise has been shown to increase the use of order information while decreasing semantic organization, and is typically associated with poor short term recall (Broadbent, 1981). Hockey and Hamilton's (1983) negative assessment of noise upon performance, however was based upon the presentation of continuous noise ranging from 80 to 100dB. Our meta-analysis data extend these research findings by examining the magnitude of effect on memory recall as noise intensity increases. The data indicate that the size of effect is related to the magnitude of the difference between noise intensity conditions. For instance the size of effect, and the decrement in performance, when moving from 85-95dB to >95dB is almost twice as large as when moving from <55dB to 85-95dB. What is more, both contrasts were significantly more detrimental to recall than when moving from <55dB to 66-78dB.

At present, the only reported effects of noise upon performance incorporated within IMPRINT is for oral communication (Allender et al., 1997; see Figure 1). By updating the stress degradation algorithms with the results of the meta-analysis, the ability to predict soldier-system performance under noise stress can be significantly improved, particularly on those cognitive tasks that include a memory recall component.

Future work will extend the current approach by analyzing related stressors and information processes already descriptively coded within the matrix.

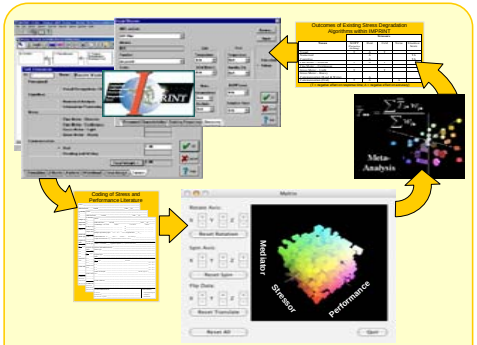


Figure 1: Evidence-based extension to IMPRINT

References

Allender, L., Salvi, L., & Promisel, D. (1997). Evaluation of human performance under diverse conditions via modeling technology. In *Proceedings of Workshop on Emerging Technologies in Human Engineering Testing and Evaluation*, June 1997. Brussels, Belgium: NATO Research Study Group.

Broadbent, D.E. (1981). The effects of moderate levels of noise on human performance. In J.V. Tobias and E.D. Schubert (Eds.), Hearing research and theory: New York: Academic Press.

Cooper, H. & Hedges, L.V. (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Hancock, P.A. & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.

Hedges, L.V. (1994). Fixed effect models. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis.* New York: Russell Sage Foundation.

Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* Academic Press.

Hockey, R. & Hamilton, P. (1983). The cognitive patterning of stress states. In G.R.J. Hockey (ed.), *Stress and fatigue in human performance* (pp.331-362). New York: John Wiley & Sons Ltd.

Lebiere, C., Bielfield, E., Archer, R., Archer, S., Allender, L., *et al.* (2002). IMPRINT/ACT-R: Integration of a task network modeling architecture with a cognitive architecture and its application to human error modeling.

Pew, R.W., & Mavor, A.S. (1998). *Modeling human and organizational behavior: Applications to military simulations.* Washington DC: National Academy Press.

APPLICATION OF FUZZY SIGNAL DETECTION THEORY TO VIGILANCE: THE EFFECT OF CRITERION SHIFTS

Shawn C. Stafford[1], James L. Szalma,[1,2] Peter A. Hancock,[1,2] & Mustapha Mouloua[1]

[1]*Department of Psychology*
[2]*Institute for Simulation and Training*
University of Central Florida

A recent advance on Signal Detection Theory (SDT) promises to enhance measurement of performance in complex real world domains. This development, Fuzzy Signal Detection Theory (FSDT), combines traditional SDT with Fuzzy Set Theory to extend signal detection analysis beyond the traditional crisp, categorical model. FSDT permits events to simultaneously be in more than one state category (e.g. signal and non-signal), so that the stimulus and response dimensions can be continuous rather than categorical. Consequently, FSDT can be employed in settings where the degree to which an event is a signal for detection may vary. This study is an initial test of application of FSDT to vigilance, a domain in which SDT has been widely applied. Results indicate that manipulations of stimulus probability impacts response bias in a fuzzy vigilance task, but that these effects differ somewhat from tasks employing traditional signal detection.

INTRODUCTION

Signal detection theory (SDT) has been extensively applied to vigilance, and research has indicated that perceptual sensitivity declines over the course of a vigil (See, Howe, Warm, & Dember, 1995). In addition, response criterion shifts to more conservative levels with time on watch (Davies & Parasuraman, 1982; See, Warm, Dember, & Howe, 1997). Note, however, that the traditional SDT model assumes that inspected events fall into discrete mutually exclusive categories (i.e. signal or non-signal). A recent advance in SDT promises to enhance measurement of performance in domains where this form of categorization is either not desirable or simply not possible. This development, Fuzzy Signal Detection Theory (FSDT; Hancock, Masalonis, & Parasuraman, 2000; Parasuraman, Masalonis, & Hancock, 2000), combines traditional SDT with Fuzzy Set Theory to extend signal detection analysis beyond the traditional crisp, categorical decision-making model. In contrast to the mutually exclusive categories of signal and non-signal in traditional SDT, FSDT allows for events to simultaneously be in more than one state category (e.g. signal *and* non-signal), so that the stimulus and response dimensions can be continuous. FSDT allows stimuli to vary in "signalness" from 0-1 by deriving mapping functions between the stimulus dimension and the 'state of the world' (characteristics of the real-world objects that are relevant dimensions for target detection). This process is referred to as 'fuzzification' (see Parasuraman et al., 2000). This permits the quantification of uncertainty inherent in the stimulus and response dimensions through the calculation of 'fuzzy' hit and false alarm rates, from which standard signal detection theory measures can be computed. Consequently, FSDT can be employed in applied settings where the degree to which an event is a signal for detection may vary. For instance, the appearance of an oblong object in airline baggage can vary in the degree to which it represents a threat. Traditional SDT would force categorization of the object into either signal or non-signal, thereby restricting the possible range of response. In contrast, FSDT captures the "fuzziness" inherent in the stimulus itself and thus uses uncertainly as a powerful source of information. The current study applied this new methodology to the vigilance problem as part of a general effort to understand the factors that control the performance, workload, and stress of tasks requiring sustained attention.

Criterion Setting in Vigilance

The FSDT methodology provides a new approach to assessing vigilance performance that incorporates aspects of the task not captured in traditional SDT. For instance, in vigilance the response criterion is especially likely to be high when signal probability is low. However, when signal probability shifts from high to low during a vigil, observers maintain a level of bias that

is more lenient than a control group that experiences a constant signal probability (see Parsons et al., 2000; Vickers & Leary, 1983). These effects occur without changes in sensitivity. It may be that application of Fuzzy Signal Detection Theory may result in performance changes that are different from those obtained using the traditional SDT model, since FSDT better captures the uncertainty inherent in the stimulus. Therefore, the current study investigated the role of 'signal probability' in controlling the criterion increment in the context of "fuzzified" stimuli. This was accomplished by varying the frequency distribution of stimulus uncertainty and evaluating the effect of changes in that distribution on the response criterion observers adopt (see Figures 1 and 2).

METHOD

Sixteen male and sixteen female undergraduate at the University of Central Florida participated in the experiment. Eight students were randomly assigned to one of four groups defined by the nature of the transition in stimulus probability. 1) Low-Low (LL): participants experienced the distribution of stimuli displayed in Figure 1, in which more 'non-signal-like' stimuli were presented than 'signal-like' stimuli; 2) High-High (HH): participants experienced the distribution of stimuli displayed in Figure 2, in which more 'signal-like' stimuli were presented than 'non-signal-like' stimuli; 3) High-Low (HL): Participants experienced the high probability distribution in session one and the low probability distribution in session 2. 4) Low-High (LH): Participants experienced the low-probability distribution in session 1 and the high probability session in session two. Note that participants were not informed of the probability distribution they would receive in either session. Prior to the first session participants engaged in a 5-minute practice vigil. The practice session was followed immediately by a two separate 20 minute vigils, each divided into four continuous five-minute periods of watch. A brief interval separated the two vigils.

The stimuli employed in this experiment consisted of two vertically oriented 1.64 cm by 20.32 cm rectangular bars separated by 1.26 cm. Each rectangle was made out of a pattern of 20,828 squares each measuring .04 cm x .04 cm. Seven different images were presented during the experiment, although participants were informed that the stimuli would vary between two extremes in color.

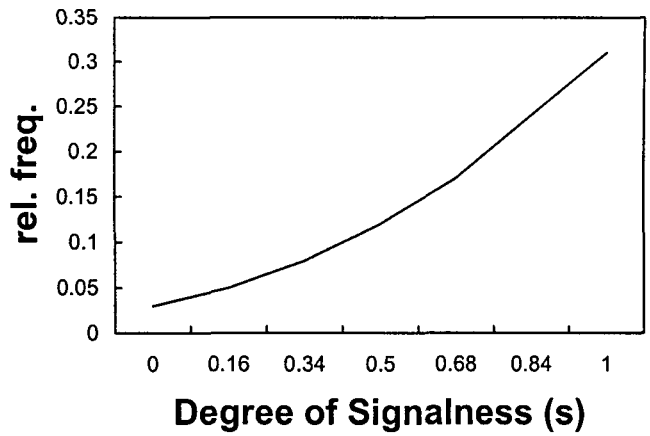


Figure 1. *Relative frequency of presentation as a function of 'signalness' (High-signal probability conditions)*

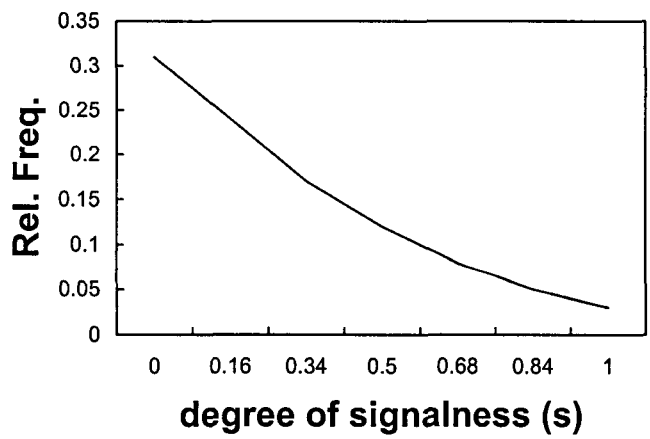


Figure 2. *Relative frequency of presentation as a function of 'signalness' (Low-signal probability conditions)*

Stimuli were presented on a Dell 1701FP monitor through an NVIDIA G-Force 3 TI-500 video card. In each image one of the two rectangles always remained a checkered pattern of squares consisting of the following two colors in equal proportions; violet squares with RGB value; Red 204, Green 0, Blue 255 and Grey squares with RGB value; Red 127, Green 127, Blue 127. The 2nd vertical rectangle consisted of checkered pattern of violet squares with RGB value; Red 204, Green 0, Blue 255, and also one of seven saturated versions of yellow squares. Using a variation of a psychophysical function developed by Indow and Stevens (1966) yellow (RGB value; Red 255, Green 255, Blue 0) was varied in its saturation level in images one thru seven for the second vertical rectangle only.

The event rate was 21 events per minute, with a stimulus duration of 200msec. Participants were

instructed to respond to stimuli by rating the degree to which each stimulus was a signal (the rectangles differed in saturation) versus a non-signal (the two bars did not differ) by pressing keys 1 through 7, with the response '7' indicating that the stimulus was definitely a signal. No other information regarding the number of possible stimulus categories or the relative frequencies of those categories was provided to the observers.

RESULTS

Responses by participants were used to compute fuzzy signal detection hit and false alarm rates using formulas obtained from Parasuraman et al. (2000), which were used to compute sensitivity (d') and response bias (c) using standard formulas (see MacMillan & Creelman, 1991). ANOVA's revealed that perceptual sensitivity did not vary as a function of stimulus probability or time on task in either session ($p>.05$ in each case).

Mean response bias scores are plotted as a function of periods of watch for each session in Figure 3. An ANOVA of response bias (the index c) revealed a significant effect for probability condition in the first session, F (3,28)=19.1, $p<.01$. Post-hoc tests using the Bonferroni correction (alpha=.05) indicated that the two conditions receiving the 'high signal probability' distribution (HH and HL) did not differ from one another in response bias in the first session, but that observers in both conditions were significantly more conservative in responding compared with observers in the two 'low signal probability' conditions (LL and LH). Note that the typical signal probability finding in vigilance, that increased signal probability decreases conservatism in responding (see Warm & Jerison, 1984; but see also Parsons et al., 2000), was not observed in the context of a fuzzy signal detection task. Instead, increased probability of 'signal-like' stimuli increased conservatism, while decreased probability of 'signal-like' stimuli induced more lenient responding.

An ANOVA of response bias in the second session indicated a significant effect for probability condition, $F(3,28)=3.08$, $p<.05$. Post-hoc tests using the Bonferroni correction indicated that observers who shifted from high to low (HL) were significantly more lenient in responding during the second session compared to those who remained on high schedule in both sessions (HH), and that observers who experienced the low probability distribution in both sessions (LL) differed significantly from those who experienced the HH condition. The bias scores of individuals who shifted from the low to the high probability distribution (LH) did not differ from those of observers in the other conditions. Thus, observers who experienced the low-probability conditions in the second session (i.e., LL and HL) were significantly more lenient than observers who experienced a high-probability distribution in both sessions.

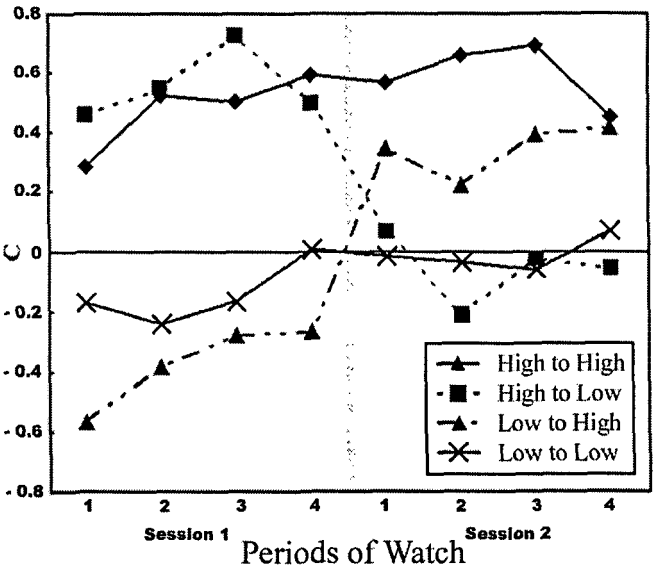


Figure 3. *Response bias as a function of 5-min periods of watch in each session.*

DISCUSSION

The current study was designed to evaluate the fuzzy signal detection theory model by investigating whether response bias effects observed in vigilance using traditional tasks and analyses would also be found using an FSDT approach. No differences in sensitivity were observed across conditions, consistent with prior findings by Parsons et al. (2000). However, in this experiment the vigilance decrement was not observed. This may be due to the moderate event rate (21 events per min) and the brief interval between the two twenty minute sessions.

In general, however, results indicated that the fuzzy detection model was sensitive to changes in response bias due to shifting stimulus probabilities. Observers in the HL condition showed the expected shift in response bias toward more lenient responding, consistent with the findings reported by Parsons et al. (2000; but see also Vickers & Leary, 1983). However, contrary to the findings of Parsons and her colleagues, observers in this study who experienced increasing signal probability (LH) did not significantly rise in conservatism relative to those in the LL condition. Moreover, Parsons et al. (2000) observed effects for time on task, while no such effects were observed herein. Specifically, Parsons and

her colleagues observed that in the decreasing condition (analogous to the HL in this experiment), responding remained at a constant, relatively unbiased level throughout the vigil, while in the increasing condition observers became more conservative over time. The differences between the findings of the current study and those of Parsons and her colleagues (2000) may be a result of the incorporation of uncertainty regarding target and non-target classification into the stimulus dimension in the FSDT model, accomplished by 'fuzzifying' the stimulus dimension (see Parasuraman et al., 2000). The resulting demands such uncertainty places upon operators, who in this experiment were compelled to rate the degree of 'signalness' rather than make discrete yes/no responses, induces bias effects different from those in more traditional experiments requiring yes/no decisions. Whether these bias effects would be observed with a fuzzy stimulus dimension but with a binary (yes/no) decision axis is a matter for future experimental work.

In addition to the differences noted above, there were also differences between studies in how probabilities were shifted. In the experiments by Parsons et al. (2000) and Vickers and Leary (1983) signal probability was shifted *within* a session, while in this study probabilities were shifted *between* sessions. This was necessary due to the fuzzification of the stimulus dimension. That is, it was not possible to define a discrete signal probability in a fuzzy task, since five of the seven stimuli were defined as having various degrees of 'signalness' and 'non-signalness.' This difference in task characteristics may also have contributed to the differential findings between this study and that of Parsons et al. (2000).

The current results reveal, however, that the explanation for the increased leniency resulting from a shift from high to low probability(contrary to the ideal observer hypothesis; see Green & Swets, 1966), may need to be reconsidered. As discussed by Parsons and her colleagues (2000; see also Parsons, 2001; Vickers & Leary, 1983), the increased leniency may be attributed to 'self-doubt' by the observers regarding the accuracy of their responses. According to this view, rather than adjust their responding to the distribution of signals, observers attribute their change in response rate during a session to their own limitations in sensitivity and adjust their responding accordingly to compensate. However, in the present experiment, probabilities associated with each stimulus remained constant and were shifted between sessions. Hence, the leniency in responding to stimuli that are primarily 'non-signal like' may be due to processes other than 'self-doubt' in response to shifting signal probabilities. The mechanism for this effect is a matter for future investigation.

Note that regardless of whether an observer experienced a change in probability distribution, response bias was more lenient when the distribution of stimuli contained a larger proportion of stimuli toward the 'non-signal' end of the continuum. That is, observers shifted their criteria to match the probability distribution they were currently experiencing, but only in the groups with the low signal probability distribution in the second session (although there was a non-significant trend for the LH group to increase in conservatism in the second session to match their cohorts in the HH group). The observation that a greater proportion of 'signal-like' stimuli induce conservatism and a smaller proportion of 'signal-like' stimuli induces leniency is inconsistent with the traditional finding in vigilance that conservatism varies inversely with signal probability. These results indicate that the conservatism in responding in a vigilance task may differ depending on whether the task is fuzzy or discrete in nature. Alternatively, it may be that criterion setting in target detection is fundamentally different in fuzzy detection tasks compared to traditional signal/noise tasks. Current empirical work suggests that fuzzy and traditional signal detection models may capture different aspects of sensitivity and response bias (Murphy, Szalma, & Hancock, 2003).

The current study has implications for real-world target detection, since in many applications the discrete categorization of events as critical events and non-critical events is not possible. For instance, in both law enforcement and combat environments, it is not unusual for signals to have characteristics of signal and non-signal, which has led to the development of guidelines for 'levels of threat' and 'levels of force.' In addition, it is not unusual for personnel in these environments to experience prolonged periods of low signal appearance, but to rapidly transition to periods of high signal probability (c.f. Warm, 1993). The results of this experiment suggest systems be designed to accommodate the shifts in response bias that occur with such transitions.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program, P.A. Hancock, *Principal Investigator*, administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr.

Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

REFERENCES

Green, D.M. Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hancock, P.A., Masalonis, A.J., & Parasuraman. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science,* **1**, (3), 207-230.

Indow, T. & Stevens, S.S. (1966). Scaling of saturation and hue. *Perception & Psychophysics*, **1**, 253-271.

Macmillan, N.A., & Creelman, C.D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Murphy, L.L., Szalma, J.L., & Hancock, P.A. (2003). Comparison of fuzzy signal detection and traditional signal detection theory approaches to performance measurement. *Proceedings of the Human Factors and Ergonomics Society*, **47**, in press.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors,* **42**, 636-659.

Parsons, K.S. (2001). *Changes in signal probability and response bias in vigilance*. Unpublished Master's Thesis, University of Cincinnati, Cincinnati, OH.

Parsons, K.S., Warm, J.S., Matthews, G., Dember, W.N., Galinsky, T.L., & Hitchcock, E.M. (2000, November). Changes in signal probability and vigilance performance. Poster presented at the meeting of the Psychonomic Society, New Orleans, LA.

See, J.E., Howe, S. R., Warm, J.S. & Dember, W.N. (1995). A meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin*, **117**, 230-249.

See, J.E., Warm, J.S., Dember, W.N., & Howe, S.R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors*, **39**, 14-29.

Vickers, D., & Leary, J.N. (1983). Criterion control in signal detection. *Human Factors*, **25**, 283-296.

Warm, J.S. (1993). Vigilance and target detection. In B.M. Huey, & C.D. Wickens, (Eds.), *Workload transition: Implications for individual and team performance* (pp. 139-170). Washington, DC: National Academy press.

Warm, J.S. & Jerison, H.J. (1984). The psychophysics of vigilance. In J.S. Warm (Ed.), *Sustained attention in human performance* (pp. 15-59). Chichester, UK: Wiley.

AN EXPERIMENTAL INVESTIGATION OF SKILL, RULE, AND KNOWLEDGE–BASED PERFORMANCE UNDER NOISE CONDITIONS

J. L. Weaver, K. A. Bradley, K. A. Burke, J. A. Helmick,
A. Greenwood-Ericksen, & P. A. Hancock
University of Central Florida
Orlando, FL

An important aspect of soldier effectiveness is the ability to perform multiple tasks in the face of stress. It has likewise been found that performance on a secondary task degrades when multiple tasks are being performed simultaneously and that the presence of stress is associated with general decrements in performance. Consequently, it is important to understand how the presence of stress impacts complex task performance in the military domain. The purpose of the present study was to investigate the effect of noise on multiple task performance, subjective workload, and subjective dimensions of stress framed within the skill – rule – and knowledge based trichotomization.

INTRODUCTION

Soldiers frequently are required to operate complex technology in stressful environments. It has been established that performance on a secondary task degrades when multiple tasks are performed simultaneously. Further, the presence of stress is associated with general decrements in performance (Hancock & Desmond, 2001). In order to alleviate these adverse effects, it is important to understand how the presence of stress impacts complex task performance. The present study is one in a series that investigates the effects of various stressors on multiple task performance. We used the Multi-attribute Task Battery (MATB; Comstock & Arnegard, 1992) composed of skill, rule, and knowledge based activities (Rasmussen, 1983) to evaluate performance variation as a function of exposure to noise.

An integral component of Rasmussen's (1983) model is attention. Two current theories of stress incorporate the importance of the delegation of limited attentional resources in response to stressors, each in a slightly different manner. Lazarus and Folkman's (1984) transactional model defines stress as *"a particular relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her well being"* (Lazarus & Folkman, 1984, p. 19). This model of stress is composed of four primary components that are theorized to proceed in sequence. Behavioral, performance, subjective, and physiological consequences are then postulated to result as outcomes of this sequential process.

Hancock and Warm (1989) propose a dynamic model of stress in which a normative zone is surrounded by a comfort zone. It is within this area that a system or individual is believed to perform at a normal target value. This zone resides between hypo and hyperstress regions indicating that a certain level of stress is necessary to operate in a normative fashion. When stress approaches either hypo or hyperstress levels, the psychological maximal adaptability zone is pressed to its limit and response eventually declines rapidly to minimal adaptability. The failure point for physiological maximal adaptability occurs further into hypo or hyperstress but represents a more serious, life-threatening danger. Performance decrements are postulated to occur as stress levels move toward hypo or hyperstress, likely due to the persons' inability to return to a normal state because of the costliness of the resources employed to remedy the problem.

In order to investigate the validity of these views of the effects of stress on multiple task performance, we used the Multi-attribute Task Battery as a test facility (MATB; Comstock & Arnegard, 1992). The MATB is a computer-based task designed to evaluate operator performance and

workload. The MATB requires the simultaneous performance of monitoring, dynamic resource management, and tracking tasks. The simultaneous performance of multiple tasks is a central feature of the MATB and it is this feature that is consistent with most operational systems and thus makes the task useful for our purposes as a research platform. Each of the MATB tasks is considered to correspond with a specific performance category in relation to Rasmussen's human performance typology. The MATB tracking task requires the manual operation of a joystick in order to keep a target in the center of a tracking window requiring little if any conscious attention and the task is therefore classified as a *skill-based* behavior. The monitoring task requires specific actions to be taken in response to each gauge in accord with an *a priori* determined set of rules, thus this task is classified as a *rule-based* behavior. The resource management task requires the activation of "pumps" in order to fill two tanks and also requires the maintenance of certain levels within them. This dynamic, *knowledge-based* task has no predetermined rules for specific operation and therefore requires conscious processing due to the ongoing problem solving nature of the task.

Hypotheses

The hypotheses regarding the current study were developed in congruence with the aforementioned notion that one must draw from a limited pool of resources in order to respond to conditions of stress and that those tasks requiring more attentional resources will degrade more in terms of performance. Specifically, it was hypothesized that the resource management task classified as knowledge-based would be shed first in response to noise (show the most disturbance under stress conditions), followed by the monitoring task classified as rule-based, and finally, the skill-based tracking task would be the last to be shed.

It was further hypothesized that NASA-TLX measures of frustration would be greater in both the stressor observation trial and the last trial of the task while effort was hypothesized to be greater in the stressor observation trial and the first trial of the task.

Finally, with regard to subjective measures of mood, it was hypothesized that such measures would indicate a decline in positive mood state over time, and that mood would be more negative in the stressor condition relative to the control condition.

METHOD

Participants

Forty-one undergraduate students (11 men and 30 women, mean age = 20.10) who volunteered to participate in this study, received extra credit for further attendance.

Apparatus

The current study used a modified version of the original MATB to investigate the effects of noise on multi-task performance. The task labels on the screen of the MATB were modified to resemble duties associated with military tank operation. An aversive artillery noise delivered through headphones at 85dB in random intervals was presented as the stressor. Performance data was collected along with data from the NASA TLX (Hart & Staveland, 1998), an instrument that measures perceived workload. Data was also collected from a demographics questionnaire, the Dundee Stress State Questionnaire (DSSQ; Matthews, Joyner, Gilliland, Campbell, Huggins, & Falconer, 1999), the Cognitive Failures Questionnaire (CFQ; Broadbent, Cooper, FitzGerald, & Parks, 1982), and the Immersive Tendencies Questionnaire (ITQ; Witmer & Singer, 1998). The questionnaires were administered via Inquisit (Millisecond, 2000), psychological data collection software for Windows, on a Dell Dimension 8200 desktop computer.

Procedure

Participants were required to fill out the demographics questionnaire, the ITQ, the CFQ, and the DSSQ upon arrival. Following the questionnaires, each task of the MATB was presented individually and the participants practiced each task for five minutes with the order of individual task presentation randomized across participants. Upon completion of each task, feedback regarding performance was provided based on normative data generated from a previous

pilot study. Participants also received feedback following a subsequent 10-minute practice session that combined all MATB tasks. Participants filled out the NASA TLX after the practice session with the assistance of the experimenter for practice purposes only.

Four observations of the combined MATB task performance were made with the occurrence of the stressor trial counterbalanced across participants to occur in either the second or third trial. Headphones were placed on the participant during each trial. During the noise stressor trial, the noise began 15 seconds after the start of the task. After each trial, participants filled out the DSSQ, NASA TLX, and a manipulation check question. The presentation order of the questionnaires was also counterbalanced across participants.

RESULTS

As a manipulation check, a one-way within subjects analysis of variance was conducted to determine whether perceived stress was higher in the noise trial. A significant main effect of trial was observed, $F(3, 105) = 3.332$, $p = .022$ ($\eta^2 = .087$). Subsequent analyses indicated that perceived stress ratings in the noise condition ($M = 56.19$) were significantly greater than those in the control observation ($M = 47.22$), the first observation trial, ($M = 47.78$), and the final observation trial 4 ($M = 42.08$).

A mixed-model 2 (condition) x 4 (trial) analysis of variance was used to determine whether performance on each MATB task declined as a function of stressor exposure across trials. Order effects were tested by including condition as a between subjects variable and trial as a within subjects variable. A significant main effect of trial was found for tracking root mean square error, $F(2.341, 86.617) = 5.062$, $p = .006$ ($\eta^2 = .120$). Subsequent analyses indicated that tracking root mean square error for observation trial 1 ($M = 101.34$) was significantly greater than in both the stressor observation trial ($M = 95.33$) and the final observation trial 4 ($M = 92.94$). Root mean square error in the control observation trial ($M = 101.06$) was significantly greater than that of both the stressor observation trial ($M = 95.33$) and the final observation trial 4 (M = 92.94). No significant main effect of condition was observed , $F(1, 37) = 1.181$, $p = .284$ ($\eta^2 = .031$).

No significant main effects of trial were observed for: monitoring correct resets, $F(3, 105) = 1.557$, $p = .204$ ($\eta^2 = .043$), monitoring incorrect resets, $F(1.714, 65.113) = 1.403$, $p = .252$ ($\eta^2 = .036$), monitoring reset time, $F(3, 99) = 1.613$, $p = .191$ ($\eta^2 = .047$), resource management tank 1 deviation, $F(3, 102) = 1.920$, $p = .131$ ($\eta^2 = .053$), or resource management tank 2 deviation, $F(3,102) = 2.567$, $p = .059$ ($\eta^2 = .070$). No significant main effects of condition were observed indicating that there were no order effects.

A one-way within subjects analysis of variance was conducted to determine whether irritation levels increased across trial. A significant main effect of trial was found, $F(3.342, 133.680) = 7.901$, $p < .0005$ ($\eta^2 = .165$). Subsequent analyses revealed that irritation levels upon arrival ($M = 3.39$) were significantly lower than those after the first observation trial 1 ($M = 2.83$), the control observation trial ($M = 2.58$), the stressor observation trial ($M = 2.49$), and the final observation trial 4 ($M = 2.90$).

Pearson's r correlation coefficients were calculated among the ITQ, the CFQ, and each of the 5 performance measures across 4 trials. The ITQ correlated positively with the CFQ, r (N = 27) = .460, p = .016 (two-tailed) and the trial 1 observation of monitoring correct reset percentage, r (N = 27) = .432, p = .024 (two-tailed). The ITQ correlated negatively with stressor observation trial resource management tank 2 deviation, r (N = 27) = -.447, p = .020 (two-tailed), control observation trial resource management tank 2 deviation, r (N = 27) = -.392, p = .043 (two-tailed), trial 4 observation resource management tank 1 deviation, r (N = 27) = -.508, p = .007 (two-tailed), and trial 4 observation resource management tank 2 deviation, r (N = 27) = -.525, p = .005 (two-tailed). Correlations among the five different performance measures within each trial were also calculated. Results indicated that none of these correlations were significant.

In order to determine whether energetic arousal, tense arousal, and hedonic tone differed across performance trials and in comparison to baseline levels of mood, a within subjects analysis of variance was conducted for each variable. Significant main effects of trial were observed for

all three of the variables, $F(4,160) = 3.093$, $p = .017$ ($\eta^2 = .072$), $F(3.328,133.111) = 6.584$, $p < .0005$ ($\eta^2 = .141$), and $F(4,160) = 5.992$, $p < .0005$ ($\eta^2 = .130$), respectively. Subsequent analyses revealed that energetic arousal scores in the second observation trial ($M = 21.61$) were significantly greater than those in the third ($M = 19.37$) and fifth ($M = 19.85$) observation trials.

Tense arousal scores on the first observation trial (baseline/score on arrival) ($M = 15.15$) were found to be significantly lower than scores on the second ($M = 18.32$), third ($M = 18.83$), fourth ($M = 18.00$), and fifth ($M = 17.39$) observation trials.

Hedonic tone scores in the first observation trial (baseline) ($M = 25.07$) were found to be significantly greater than those in the second ($M = 23.30$), third ($M = 21.90$), and fourth ($M = 22.56$) observation trials, indicating a more pleasant mood in the first trial.

In order to determine whether energetic arousal, tense arousal, and hedonic tone differed in the first observation trial (baseline) from subsequent performance trials (including the stressor vs. control conditions), a within subjects analysis of variance was conducted for each variable. Significant main effects of trial were observed for all three variables, $F(4,160) = 3.334$, $p = .012$ ($\eta^2 = .077$), $F(4,160) = 6.852$, $p < .0005$ ($\eta^2 = .146$), and $F(4,160) = 5.834$, $p < .0005$ ($\eta^2 = .127$), respectively.

Subsequent analyses revealed that energetic arousal scores in the first observation trial (baseline) ($M = 20.78$) were significantly greater than those in the control observation trial ($M = 19.27$). Subsequent analyses also indicated that tense arousal scores in the first observation trial ($M = 15.15$) were significantly lower than those in the second ($M = 18.32$), experimental ($M = 18.98$), control ($M = 17.85$), and final observation ($M = 17.39$) trials.

Hedonic tone scores in the first observation trial ($M = 25.07$) were found to be significantly higher than those in the second ($M = 23.29$), experimental ($M = 22.02$), and control observation ($M = 22.44$) trials, indicating a more pleasant mood in the first observation trial. There was no difference between the stressor or control conditions for any of these three variables.

A within subjects analysis of variance was used to determine whether effort was greatest in the first observation trial. A significant main effect of trial was found, $F(2.328, 90.798) = 5.079$, $p = .006$ ($\eta 2 = .120$). Subsequent analysis revealed that, in accord with the initial hypothesis, effort levels in the first observation trial ($M = 26.30$) were significantly greater than in the second observation ($M = 15.62$), the third observation, ($M = 16.98$), and the fourth observation, ($M = 18.18$). A within subjects analysis of variance was also used to determine whether frustration was greatest in the last trial. A significant main effect of trial was found, $F(3, 120) = 4.808$, $p = .003$ ($\eta 2 = .107$). However, subsequent analyses indicated that frustration was not highest in the last trial, but lower in trial one relative to all others.

A paired samples t-test was performed in order to determine whether effort was greatest in the noise trial. In order to control for time, effort scores in the first, last, and control conditions were averaged and compared to the noise condition mean score. Contrary to the hypothesis, effort in all other trials ($M = 20.48$, $SD = 13.51$) was significantly higher than in the noise condition ($M = 15.64$, $SD = 14.75$), $t(39) = -2.451$, $p = .010$ (one-tailed).

A paired samples t-test was also performed in order to determine whether frustration was greatest in the noise trial. In order to control for time, frustration scores in the first, last, and control conditions were averaged and compared to the noise condition mean score. Contrary to the hypothesis, frustration in the noise condition ($M = 30.42$, $SD = 23.09$) was not significantly higher than all other trials ($M = 25.63$, $SD = 17.68$), $t(40) = 1.319$, $p = .0975$ (one-tailed).

DISCUSSION

The purpose of the current study was to determine the degree that aspects of complex task performance would degrade under conditions of stress (i.e., noise). It was also a critical concern to assess the relationship between components of task performance and several affective measures (i.e., DSSQ and NASA-TLX), as well as subjective measures reflective of cognitive style (i.e., CFQ and ITQ).

In general, the performance findings did not support our original degradation hypotheses. Although the manipulation check regarding

perceived stress showed that the experimental condition differed significantly from the other trials, there was only one significant performance finding. One possible explanation for this finding was that the task itself was so stressful that the addition of the noise failed to result in any meaningful additional difference. In order to test this possibility, the question regarding subjective irritation from the DSSQ was compared across the four task performance conditions with the response given by subjects upon their arrival. Interestingly, this comparison indicated that subjects were significantly less irritated upon arrival in comparison to all task performance trials. Thus, perhaps the task itself was consistently stressful enough in a performance context that performance was only minimally influenced.

The one significant performance finding related to tracking performance. However, tracking performance was *improved* by noise exposure. Although this was not an *a priori* hypothesized finding within the context of the current study, the findings of stress researchers such as Poulton (1978) provide insight regarding this result. It is likely that participants were stimulated by the noise in this condition because they had been performing the task for at least 45 minutes. Thus, the noise countered subjective fatigue possibly being experienced by the participants at this time. Further analyses of our remaining data will likely shed further light on this finding. It is hoped that the findings of this research program will ultimately contribute to our understanding of the effects of various stressors on different types of tasks in the military domain. The current study serves to illustrate that stressor effects are often non-intuitive and in fact often counter to what is hypothesized. Most important is to consider the utility of the MATB as truly illustrative of the components of Rasmussen's typology. Future efforts will be devoted to this and other issues (e.g., choice and manipulation of stressors) related to the continuation of this research program.

ACKNOWLEDGEMENTS

This research was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, P.A. Hancock, Principal Investigator. The views expressed are those of the authors and do not necessarily represent those of the United States Army.

REFERENCES

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The cognitive failures uestionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21,* 1-16.

Comstock, J. L., & Arnegard, R. J. (1992). *The Multi-attribute Task Battery for human operator workload and strategic behavior research.* Technical Report 104174. Hampton, VA: NASA Langley Research Center.

Hancock, P. A., & Desmond, P. A. (Eds.). (2001). *Stress, workload, and fatigue.* Mahwah, NJ: Lawrence Erlbaum.

Hancock, P. A., & Warm. J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519-537.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Millisecond Software (2002). Inquisit 1.XX [Computer software]. Seattle, WA: Millisecond Software.

Lazarus, R. S., & Folkman, S. (1984). *Stress appraisal and coping.* New York: Springer.

Matthews, G. Joyner, L., Gilliland, K., Campbell, S. Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (vol 7, pp. 335-350).

Poulton, E. C. (1978). A new look at the effects of noise: A rejoinder. *Psychological Bulletin, 85,* 1068-1079.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC 13,* 257-266.

Witmer, B. G., Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225-240.

MODERATING FACTORS IN VISUAL SEARCH: THE ROLE OF ECOLOGICAL-VALIDITY

Derek D. Diaz, Valerie K. Sims, Peter A. Hancock,
Hana S. Smith, Linda Upham Ellis, Bryan Clark, & David Sushil
University of Central Florida

Participants completed a visual search task that varied along three dimensions: 1) stimulus type: ecologically valid humans or ecologically neutral objects, 2) stimulus behavior: static, dynamic or random, and 3) number of targets: one or two. Search was faster when ecologically valid stimuli were used, and when these moved in a predictable fashion. Certain types of visual searches require search memory, or the ability to mentally tag previously examined objects. This search memory is optimized when ecologically valid stimuli are used. Design of visual search interfaces should take into account both the task and the stimuli.

INTRODUCTION

In an effort to limit confounds and establish reliable sets of stimuli, most visual search experiments have used simple alphanumeric or esoteric symbols (c.f., Baddeley, 1990), and relatively few have taken complex environmental stimuli into account (c.f., Wolfe, J.M., Oliva, A., Butcher, S., & Arsenio, H., 2002). Therefore, most of the studies on visual search examine performance of visual tasks of stimuli that were not available during the evolutionary development of our visual system (Baddeley, 1990). Unlike the stimuli in these laboratory studies, stimuli in the natural world provide many attentional cues, including biological motion. We refer to these types of natural stimuli as those which obey ecologically-valid principles. The purpose of the present study is to examine visual search with ecologically valid stimuli. Specifically, we examine the possibility that the cognitive processes underlying visual search are highly dependent on stimulus type.

The Case for Search Memory

Eye-movements during visual search of a complicated scene are not random. Gaze patterns across environmental information can be systemically influenced by internal (cognitive) effects (Treisman & Gelade, 1980; Kristjansso, 2000), external characteristics of the stimuli (Downing & Picker, 1985), and their interactions. Although these two sources of variance are not always independent, the effects of external influences are arguably better understood than effects from internal sources. One theory regarding the internal cognitive aspects of visual search postulates that it is guided by a tagging process which inhibits the return of visual attention to previously seen locations (Klein 1988). This facilitative effect of "search-memory" was confirmed in a recent study by Gibson, Li, Skow, Brown, and Cooke (2000). In this work, participants were asked to verify whether an array contained two identical targets or a single target. One condition allowed the use of search-memory and another did not. Both conditions showed moving stimuli. However, in the search memory facilitating condition, movement was continuous, whereas in the memory prohibitive condition, targets and distractors were repositioned every 107 msec, making it impossible to keep track of a given target. The authors concluded that search-memory did facilitate visual search because performance (reaction time and accuracy) was better for conditions which could use search-memory.

Present Study

This study investigates whether the ecological validity of the visual scene (search-space) may be one among other possible mediators in the efficiency of search memory. Our theory is that search-memory is more effective when the search-space has a high degree of ecological-validity. To test this theory, we have conducted an experiment which varied both the ecological validity of the search-space and the participant's ability to use search memory. The ecological validity of the search-space was manipulated with two variables: stimuli-shape (either people or symbols) and stimuli-behavior (movement, no movement, or random displacement). The objects with high ecological-validity were people and the stimuli without ecological significance (neutral objects) were iconic shapes. Thus, search-spaces with different levels of ecological validity were created using combinations of these two variables. The moving-people combination was expected to be a powerful affordance, and thus represent the condition with the most ecological validity. The ability to use search memory was manipulated by randomly repositioning the search-space items on half of the trials (random-displacement).

By presenting either 1 or 2 identical targets on a single trial, we planned to replicate the results of Gibson et al. (2000), and again confirm that search-memory facilitates performance in the double-target detection task. Thus, we hypothesize a dissociation between the random-displacement and no-movement conditions. Furthermore, we expect that performance will be better for search-spaces with ecologically valid conditions (e.g., moving people stimuli) than non-ecologically valid conditions (moving symbols).

METHOD

Participants

Twenty-one psychology undergraduates from the University of Central Florida received extra credit for participating in the study. The average age of the participants was 20.2 years. All reported normal or corrected-to-normal visual acuity and were screened for common color deficiencies.

Apparatus

A Dell Dimension 8200 computer (Intel Pentium 4-2.4 GHz microprocessor, 512 MB synchronous dynamic RAM, Nvidia Gforce 4 Titanium graphics card) running Microsoft Windows XP was used to conduct the experiment. Stimuli were presented on a 17 inch SVGA CRT video monitor (75 Hz vertical refresh rate, 1024 X 768, 16-bit color). Data collection and stimuli presentation were accomplished using Inquisit software (2002, Millisecond Software, Seattle, WA).

Stimuli

Two sets of stimuli were used in the study: ecologically-valid (EVS) and ecologically-neutral (ENS). Both sets included 15 items made up of targets and distractor objects. Images for the EVS set were created from three-dimensional human models. Objects in this group were soldiers, and the target was a more lithe soldier. The ENS set was made up of meaningless symbols, and targets were mirror images of the distractors. These shapes were similar to those used in previous studies (e.g., Gibson et al., 2000; Wang, Cavanagh, & Green, 1994). However, the coloring of the ENS stimuli was taken from the hue and luminance distribution of the EVS stimuli. Thus, ENS distractors contained the same colors as EVS distractors and ENS targets contained the same colors as EVS targets. Both types of stimuli were 2.0 cm high and 1.0 cm wide, and subtended a vertically and horizontally visual angle of 1.56°. Grayscale representations of the stimuli are shown in Figure 1.

On any given trial, stimuli presented behaved in one of three possible ways. After appearing, the objects either did not move (no-movement), continuously moved in a random linear direction (movement), or disappeared and reappeared in a random location every 250 ms (random).

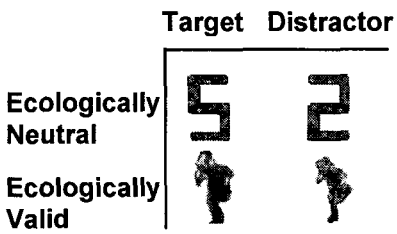


Figure 1. Stimuli used in the study (stimuli actually appeared in color).

Procedure

Twelve conditions were created by combining three factors: stimuli-class (EVS-people or ENS-shapes), stimulus-behavior (movement, no-movement, or random), and target-number (one or two). After receiving instructions for the experiment, participants were guided through a practice session that lasted for approximately 5 minutes. Participants were seated 23 inches from the display, and asked to fixate on a cross positioned in the center of the display. This fixation point was replaced by a search-set after 500 ms. The search-set was randomly selected from the 12 possible conditions. Half of the trials contained 1 target and the other half contained two. The task was to respond by pressing a button labeled 1 or 2 after deciding how many targets were present in the display. Participants were instructed to respond using their index finger as quickly and accurately as possible. To ensure that participants did not misplace their finger at the start of each trial, a tactile cue indicated a neutral position between the two response buttons. Stimuli for a single trial lasted for a maximum of 15 s. The display was replaced by a black mask immediately after a response was given. No performance feedback was given. An inter-trial interval of 2 seconds was used. Each of the 12 conditions was presented to the participant 24 times to create a total of 288 trials. These trials were administered across 6 blocks of 48 trials. For every block, each participant received a randomly generated sequence of conditions. Blocks were separated by short breaks of up to 4 minutes.

RESULTS AND DISCUSSION

A within-subjects 2 (stimulus type: people or shapes) x 3 (stimulus-behavior: static, dynamic or random) x 2(target-number: one or two) ANOVA was calculated with reaction time as the dependent variable. Experiment block did not account for a significant amount of variance, and was not included in the analyses presented. Reaction time results were derived only from trials with correct responses, and any response made after 15 seconds was excluded from the analysis.

The Search Memory Hypothesis

A main effect of stimulus-behavior was found, $F(2, 36) = 159.81$, $p<.01$, *eta-squared* = .90. As indicated by the scatter plots in Figure 2 below, there is a clear separation (M=3793.99 msec) in the latency of responses under the random-displacement (M=6615.10 msec, SD=309.77 msec) and no-movement conditions (M=2821.11 msec, SD=150.90 msec) for both types of stimuli. These results are similar to that reported by Gibson et al. (2000) and provide added support for the hypothesis that search-memory facilitates visual search.

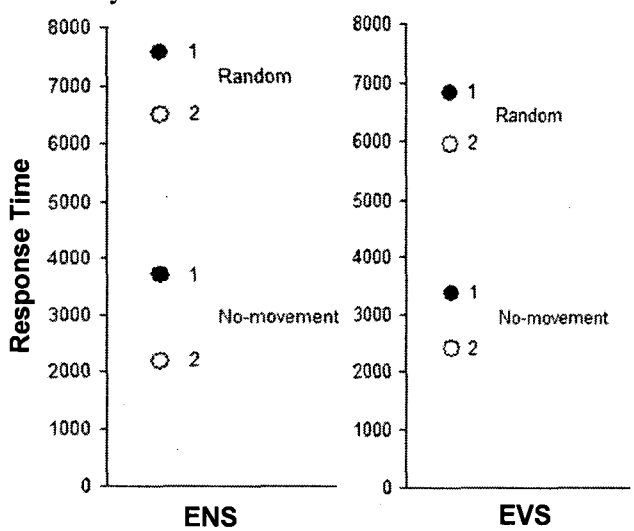


Figure 2. Response time to Ecologically Valid Stimuli (EVS) and Ecologically Neutral Stimuli (ENS) during random-displacement and no-movement conditions.

Stimulus-type (EVS versus ENS)

The analysis also produced a significant main effect of stimulus-type, $F(1, 18) = 40.34$, $p<.01$, *eta-squared* = .69. Participants were faster when making correct responses to EVS (M=4094.74 msec, SD=184.67 msec) than to the ENS (M=4583.99, SD=224.98 msec). A two-way interaction between stimulus-type and stimulus-behavior also was significant, $F(2, 36) = 12.52$, $p<.01$, *eta-squared* = .41. As hypothesized, reaction time for EVS (M=3351 msec, SD=1193 msec) was significantly faster than for ENS (M=4043 msec, SD=1536) under the "movement" stimulus-behavior condition. Post hoc analyses indicated that EVS had significantly shorter response time than ENS only under the "movement" stimulus-behavior condition, $t(41)=4.80$, $p < 0.05$. Thus, as illustrated in Figure 3 below, when the stimuli presented were people moving naturally, participants responded more quickly than when presented with symbols moving in the same way. In addition, the difference between the means for EVS (M= 2882 msec, SD= 965.68 msec) and ENS (M= 2940 msec, SD= 1079.52 msec) under the no-movement stimulus-behavior condition was not significant, suggesting that ecological validity requires that the stimulus and its movement must "agree." That is, the visual system responds more quickly when a stimulus moves in a way that is predictable in the world.

Search-Memory and Stimulus-Type

The three-way interaction from the omnibus ANOVA also was significant, $F(2, 36) = 5.09$, $p<.01$, *eta-squared* = .22. This analysis shows that the effects of ecologically valid stimuli and movement were seen primarily in the 2-target condition, when search-memory could be used. As in the other analyses, with 2 targets, participants responded significantly more rapidly to EVS than ENS under the "movement" stimulus-behavior condition. This did not occur for the "no-movement" stimulus-behavior condition during searches for 2-targets. Again, the results suggest that optimal visual search memory is optimized when stimuli are ecologically valid in two ways--a familiar shape moving in a predictable manner.

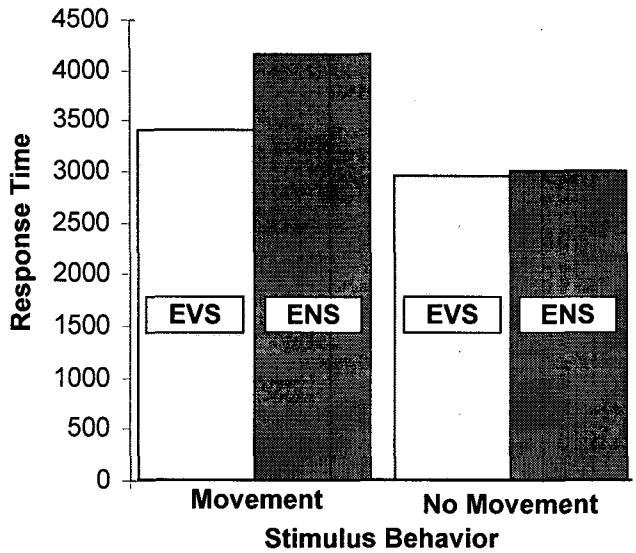


Figure 3. Response time to Ecologically Valid Stimuli (EVS) and Ecologically Neutral Stimuli (ENS) when the stimuli were dynamically moving or not moving at all.

CONCLUSIONS

The data presented here further support the argument that visual searches are conducted more quickly when one can utilize an internal search memory process. Further, the results demonstrate that search memory itself can be optimized by characteristics of the stimuli. Ecologically valid stimuli, or those real world objects that behave in real world ways, appear to be processed more quickly. Thus, efficiency of a visual search can be predicted using a two step approach. First, one must assess whether the task lends itself to using search memory. If search memory is indicated, then the nature of the stimuli must be assessed. Ecologically valid stimuli, in terms of both form and movement, will lead to the greatest efficiency.
Visual search in virtual and augmented environments may be made much more efficient if the stimuli and their movement are designed to correspond to the strengths of the human visual system. In this case, humans appear to have evolved to recognize moving humans more

quickly than moving objects. Further research should address whether this effect of ecological validity is limited to human form stimuli. A human form may require fewer resources to represent in working memory, as these stimuli may be retrieved from long term stores. Perhaps other highly familiar stimuli that obey predictable movement patterns, would yield similar results. Additionally, further work needs to tease apart the variables of ecological form and ecological movement. If both of these aspects of the stimulus are not necessary, design of search environments may be much simpler.

REFERENCES

Baddeley, A. D. (1990). Human Memory: Theory and Practice, Revised ed. Needham Heights, MA: Allyn and Bacon.

Downing, C. J. & Picker, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds). Attention and Performance XI. Hillsdale, NJ; Erlbaum, 171-187.

Gibson, B. S., Li, L., Skow, E., Brown, K., & Cooke, L. (2000). Searching for one versus two identical targets: When visual search has a memory. Psychological Science, 11, 324-327.

Horowitz, T.S. & Wolfe, J. M. (1998). Visual Search has no memory. Nature, 357, 575-577.

Inquisit 1.0.32.0 [Computer software].(2002). Seattle, WA: Millisecond Software.

Klein R. & MacInnes, W.J. (1999). Inhibition of return is a foraging facilitator in visual search. Psychological Science, 10, 346-352.

Klein R. (1988). Inhibitory tagging system facilitates visual search. Nature, 334, 430-431.

Kristjansson, A. (2000). In search of Remembrance: Evidence for memory in visual search. Psychological Science, 11(4), 328-332

Treisman, A., & Gelade, G. (1988). A feature integration theory of attention. Cognitive Psychology, 12, 97-136.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. Psychonomic Bulletin & Review 1, 202-238.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. Perception and Psychophysics, 56, 495-500.

Wolfe, J. M. (1996). Visual search. In H. Pashler (Ed.), Attention, . London, UK: University College London Press.

Wolfe, J.M., Oliva, A., Butcher, S., & Arsenio, H. An unbinding problem: the desintegration of visible, previously attended objects does not attract attention. Journal of Vision, 2, 256-271. (2002).

AN INVESTIGATION OF INDIVIDUAL DIFFERENCES IN RELATION TO WARNING COMPLIANCE AND USER AFFECT: PRELIMINARY FINDINGS

Jeanne L. Weaver, Jessica A. Helmick, & Kelly A. Burke
University of Central Florida
Orlando, Florida

Consumers frequently encounter warning labels and safety precautions on the products they purchase, whether it is at home or in the workplace. Studies on the physical aspects of warning labels have been extensively considered, such as variability in font size, weight, size contrast, etc. Although it is important to consider the physical characteristics of the warning label, individual characteristics as they pertain to consumers are also important to consider. For example, Weaver, et al. found that health orientation and sensation seeking were significantly related to behavioral compliance to warnings. The current study investigated the extent to which participant characteristics would relate to warning compliance when exposed to a noise stressor.

INTRODUCTION

Much of the research that has been conducted regarding warnings and compliance has focused on aspects of the warning itself (e.g., type of signal word, color, placement, etc.) and/or person variables that influence behavioral compliance to the warning (e.g., age, gender, hazard perception, etc.). However, some of this research has yielded equivocal results. Consequently, Weaver, Gerber, Hancock, and Ganey (2003) explored the role of sensation seeking and health orientation as individual difference variables, in addition to the effects of signal word presentation on behavioral compliance to a health warning. Specifically, they tested the proposition that one's health orientation and level of sensation seeking would relate significantly to compliance with regard to warnings of *different strengths*. The results of the study provided support for these hypotheses in that those who were higher in sensation seeking required a stronger signal word to gain compliance. Also of interest was the finding that individuals low in health orientation were less likely to comply in the "no warning" condition than those with higher health orientation. These data serve to support the general conception that individual difference factors such as health orientation and sensation seeking require consideration in evaluating the nature of warning compliance. In addition to these findings confirming that individual differences might play a significant role in determining the extent to which persons comply with warnings, they also indicate a potential explanation for past equivocal findings in warning research regarding the manipulation of signal word strength.

Weaver et al. (2003) proposed in their discussion that arousal might be the mechanism responsible for the results of their study with regard to sensation seeking differences in compliance. Specifically, they speculated that persons who differ with regard to their sensation seeking might require warnings conveying different levels of hazard in order to raise arousal level to a point at which compliance seems necessary. Arousal state might also explain other findings in which those with more experience with a product are typically shown to exhibit lower compliance to warnings presented with that product (Zeitlin, 1994). Indeed we propose that a certain degree of arousal is necessitated to insure compliance. In Zeitlin's work, for example, anxiety reduction was the main reason given for compliance with safety instructions for users inexperienced with a chainsaw. Consequently, the study reported here acts as a follow-up to Weaver et al. and seeks specifically to consider the role of arousal in relation to individual differences and compliance to warnings of varying strengths.

In addition, it was also of interest within the current study to determine the degree that compliance might be influenced by the presence of

a stressor. Given that products are often utilized by consumers under potentially stressful conditions, the current study also sought to determine the influence on compliance when participants were placed in stress (noise) vs. non-stress conditions. Finally, the study reported here also extends the work described in Weaver et al. (2003) by increasing the possible number of compliance behaviors and by utilizing a unique methodology (a wood-working task) in order to study the above variables.

It was hypothesized 1) that there would be a main effect of warning condition, with compliance highest in the "Danger" condition and lowest in the control "Notice" condition, 2) that there would be an interaction of sensation seeking with level of warning condition such that those who were higher in sensation seeking would require a stronger signal word to gain compliance than low sensation seekers, 3) that there would be a main effect of sensation seeking with low sensation seekers more likely to comply overall, 4) that high sensation seekers would report lower arousal upon arrival to the experiment, 5) that sub-scores on the NASA-TLX would differ across stressor condition, as well as compliance being lower in the stressor condition. Specifically, it was hypothesized that participants would report higher mental demand, lower self-rated performance, and higher effort and frustration in the noise condition. The following describes the ongoing study and provides results analyzed for the first 54 of 120 participants to be observed

METHOD

Participants

Fifty-four (of the intended 120 participants) female University of Central Florida undergraduate students volunteered for this experiment in exchange for extra credit in an undergraduate psychology course. The experiment was announced under the guise of a team performance exercise in order not to compromise the true nature of the experiment. Participants were recruited via the Psychology Department Research Participation web site. Also, participants were recruited via a sign-up sheet obtained directly from undergraduate psychology courses. All participants were treated in accordance with the "Ethical Principles of Psychologist and Code of Conduct" (American Psychological Association, 1992).

Materials

The questionnaires administered followed the same approach taken by the Weaver et al. (2003) experiment regarding warning compliance with the addition of a subjective measurement of workload and the team performance questions. The participants completed the following questionnaires: NASA-TLX (Hart & Staveland, 1988), Zuckerman's Sensation Seeking Scale-Form V (1994), an exit survey adapted from Braun's (1993) study on Color Product Warnings and Behavioral Compliance, Otsubo's (Otsubo, 1998) behavioral study of warning labels and consumer products, a demographic questionnaire, and several questions about team performance. These questions elicited information about the participant's team performance experience and was administered solely for the purpose of maintaining the guise of the experiment's focus on team performance. For more information regarding these questionnaires see Weaver et al. All of the questionnaires, with the exception of the demographic questionnaire, were administered via the software program Inquisit (Millisecond, 2002) on a Dell Dimension 8200 desktop computer.

In order to observe the participant's compliance behaviors, a small home surveillance camera was installed in the tool work station hidden from the participant's view. The camera was connected to a small color video display terminal (VDT) with a 12" diagonal viewing area which was located in the room adjacent to the tool work station. The equipment used for this experiment was four small pieces of wood (6" X 4" X 2"), sandpaper, sanding block, paintbrushes, paint, paint sponges, corn oil. The compliance items included a paper mask, goggles, a protective jacket, rubber gloves, eyewash, and eyewash cups. Other tools available on the tool station were a hammer, nails, glue, and two additional paint cans, even though their use was not necessary or referred to in the instructions. The noise stressor was manipulated using an audio cassette recording of a table saw cutting wood. The recorded stressor was presented in random intervals

at 85 decibels for the duration of the experimental phase. The noise level was calibrated with a sound pressure meter prior to the start of each experimental day to ensure that it was at 85 decibels. A Pioneer dual cassette player/recorder and two Bose speakers were used to present the noise stress. The cassette player was hidden from view of the participants so that it would be believed that the wood cutting was taking place in the next room by the other "team member."

Procedure

Participants were randomly assigned to one of eight conditions based on the signal word condition and the noise stress. To increase the realism of the team performance task, a confederate, assuming the role of another participant, arrived at the same time as the actual participant. The two "participants" were briefed about the "Team Performance" experiment and then asked to read and sign an informed consent form. The participant was separated from the confederate and seated in front of the computer to complete the first set of questionnaires. After completing the first set of questionnaires, the participant was directed into the tool work station where the experimenter asked several familiarity questions (to determine wood working experience). Next, she was asked to first fill out the demographic questionnaire, read the demonstration instructions, and then begin the wood working task. Immediately after the participant received these instructions, the experimenter gave the same instructions and familiarity questions to the confederate. The participant was observed by the experimenter via the VDT and scored on their compliance behaviors. In the stress condition, the experimenter turned on the cassette player with the noise stress as soon as the participant began reading the demonstration instructions. After completing three of the five task instructions, the experimenter instructed the participant to stop the task, remove all of the protective equipment, and begin the final set of questionnaires on the computer. The experimenter also instructed the confederate to stop the wood cutting task and then turned off the cassette player (applies to stress condition only). After the participant completed the final set of questionnaires, she was debriefed in the form of both verbal and written statements.

RESULTS

The dependent measures of interest were the behavioral compliance scores (the sum of the number of behaviors performed by each participant in each condition), Workload (as measured by the NASA-TLX), and Sensation Seeking (as measured by Zuckerman's Sensation Seeking Scale-Form V). For the analyses to be described, the between-subjects factors were the stress condition with two levels (noise, no noise) and the degree of threat of the warning with four levels (control, notice, warning, and danger). All analyses were based on an alpha level of .05.

Manipulation Check

As a manipulation check, an independent samples *t*-test was conducted to determine whether perceived stress was higher in the noise stress condition. Results yielded no statistically significant difference for this analysis, $t(52) = 1.18, p = .12$, (one-tailed) indicating that participants did not report feeling more stressed in the noise condition. Consequently, stressor condition was not included in subsequent factorial analyses.

Behavioral Compliance

Behavioral compliance was defined as the donning of protective equipment (mask, gloves, goggles, and jacket) and filling eyewash cups which was measured on a 5-point scale with "5" indicating that all protective equipment was worn and that the eyewash cups were filled, and "0" indicating that none of the protective equipment was worn and the eyewash cups were not filled. The behavioral compliance score was obtained by summing the number of behaviors performed by each participant in each condition.

A between-subjects 2 (low SSS vs. high SSS) X 4 (control, notice, warning, and danger) ANOVA on behavioral compliance revealed a main effect of threat $F(3, 44) = 20.321, p < .0005$ indicating that compliance was influenced by threat level. However, subsequent tests indicated that the

control condition differed only from the other three threat conditions. Participants were less likely to comply with the message in the Control condition (M = .73) than the Notice (M = 4.8), Warning (M = 4.28), or Danger (M = 4.07) conditions.

There was no main effect of sensation seeking nor did the interaction between threat level and sensation seeking reach significance. The values for these effects were $F(1, 44) = .108$, $p > .05$, and $F(3, 44) = .612$, $p > .05$, respectively. Thus, neither of these hypotheses was supported.

An independent samples t-test was conducted to determine if there was a difference in behavioral compliance between stress conditions. Contrary to the hypothesis, no significant difference emerged, $t(52) = -.83$, $p = .20$ (one-tailed).

NASA-TLX

As noted in the end of the introduction, four *a priori* hypotheses were predicted with regard to the independent rating scales for Mental Demand, Performance, Effort, and Frustration on the NASA-TLX. In order to test these hypotheses, four independent samples t-tests were conducted to identify whether the noise stress affected the participant's subjective workload as measured by these rating scales. Results indicated significant differences across stress conditions for both performance and frustration $t(51) = 2.080$, $p < .05$ and $t(51) = -2.314$, $p < .05$ (both one-tailed). Although participants in the noise condition rated their performance as lower than in no noise condition, reported frustration was higher in the no noise condition. There were no significant differences for mental demand or effort ($p > .05$).

Arousal

In order to test the hypothesis that low sensation seeking persons would report lower arousal, an independent groups t-test was conducted. Results indicated no significant difference between the groups $t(50) = .66$, $p > .05$ (one-tailed).

DISCUSSION

The purpose of the current study is to act as a follow-up to the study by Weaver et al. (2003). It was specifically of interest to consider the role of individual differences and compliance to warnings of varying strengths. In addition, the current study is also investigating behavioral compliance when participants are placed in stress (noise) vs. non-stress conditions. Finally, this study extends the work described in Weaver et al. (2003) by increasing the possible number of compliance behaviors. The study also utilizes a unique methodology (a wood-working task) with relatively higher realism than some other warnings methodologies.

It is important to note that the results described here represent only a portion of the data that is currently being collected (i.e., 54 of 120 participants), thus, the results reported reflect only those observations. It is our belief that null results related to the manipulation check question, for example, will likely be significant with stronger power.

Although there was not a significant difference in compliance across stress conditions, there was a significant finding for compliance as a function of threat, with the control condition being associated with the lowest levels of compliance. It is believed that this finding will also become more defined with the addition of more participants. That is, we would then expect to see differences in compliance between other levels of threat as well. In conclusion, it is anticipated that the current study will replicate the findings of Weaver et al. (2003) and others in addition to making a contribution via the novel aspects of the investigation (e.g., the new methodology for the investigation of warnings) and perhaps the presence of a stressor with regard to compliance.

REFERENCES

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47, 1597-1611.

Braun, C.C. (1993). *Perceptions as hazard as a function of color and signal word in product warnings.* Unpublished Master's Thesis, University of Central Florida, Orlando, FL.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload, (pp. 139-183). Amsterdam: North-Holland.*

Millisecond Software (2002). Inquisit 1.XX [Computer software]. Seattle, WA: Millisecond Software.

Otsubo, S.M. (1998). A behavioral study of warning labels for consumer products: Perceived danger and use of pictographs. *Proceedings of the Human Factors Society 32nd Annual Meeting*, 536-540. Santa Monica, CA: Human Factors and Ergonomics Society.

Weaver, J.L., Gerber, T.N., Hancock, P.A., & Ganey, H.C.N. (2003). Individual differences in behavioral compliance to warnings representing varying degrees of threat. *International Journal of Occupational Safety and Ergonomics*, *9*(2), 147-158.

Zeitlin, L.R. (1994). Failure to follow safety instructions: Faulty communication or risky decisions? *Human Factors*, *36*(1), 172-181.

Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking.* New York: Cambridge University Press.

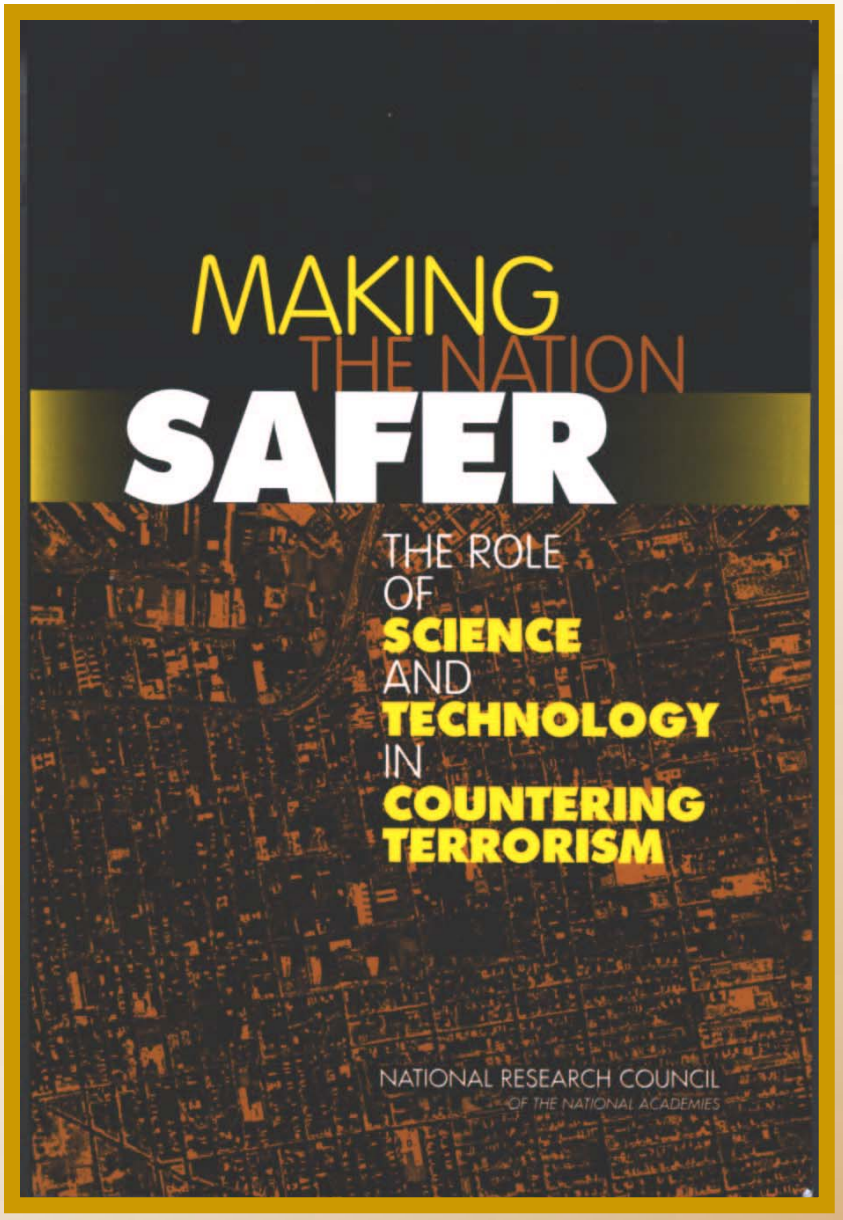
Application of Psychology to Homeland Security





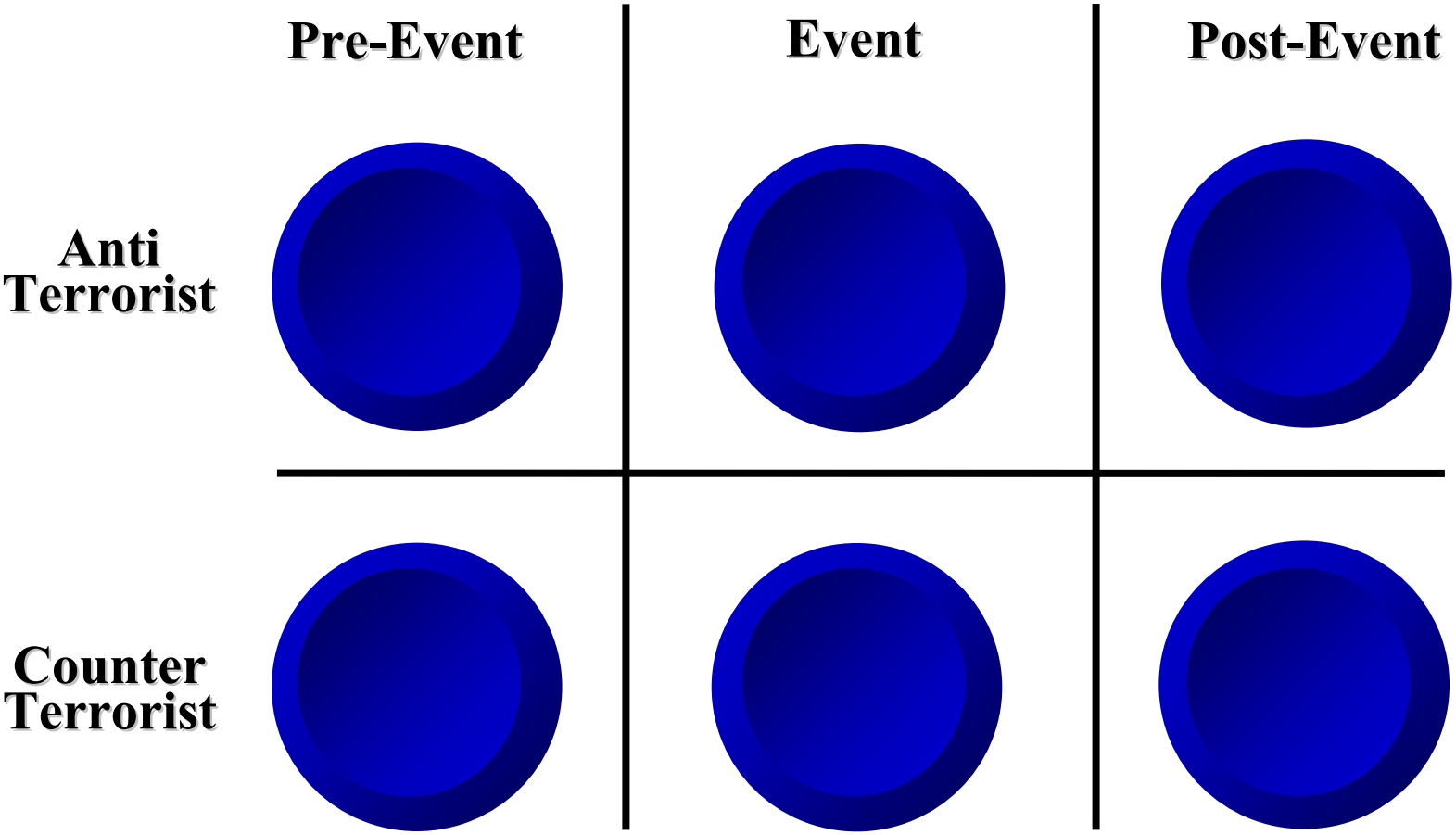
Peter A Hancock, Ph.D.
University of Central Florida

Southeastern Psychological Association, March 26-29, 2003.



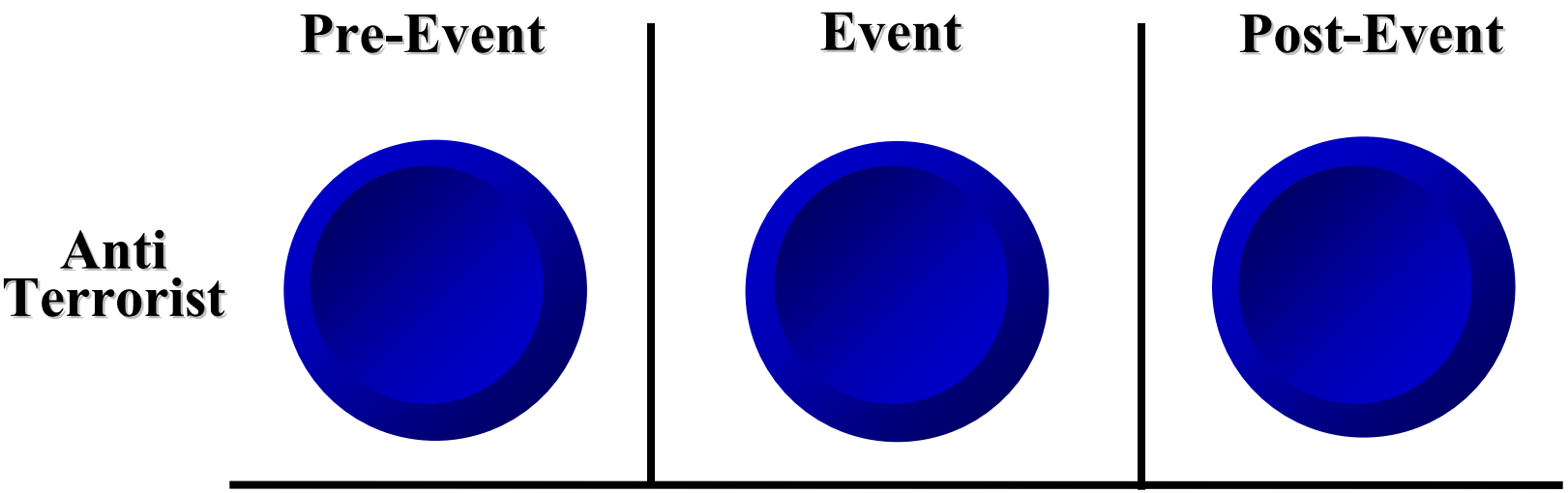
Report of the National Academies on the Issues of Homeland Security

Human Factors Contributions to Homeland Defense

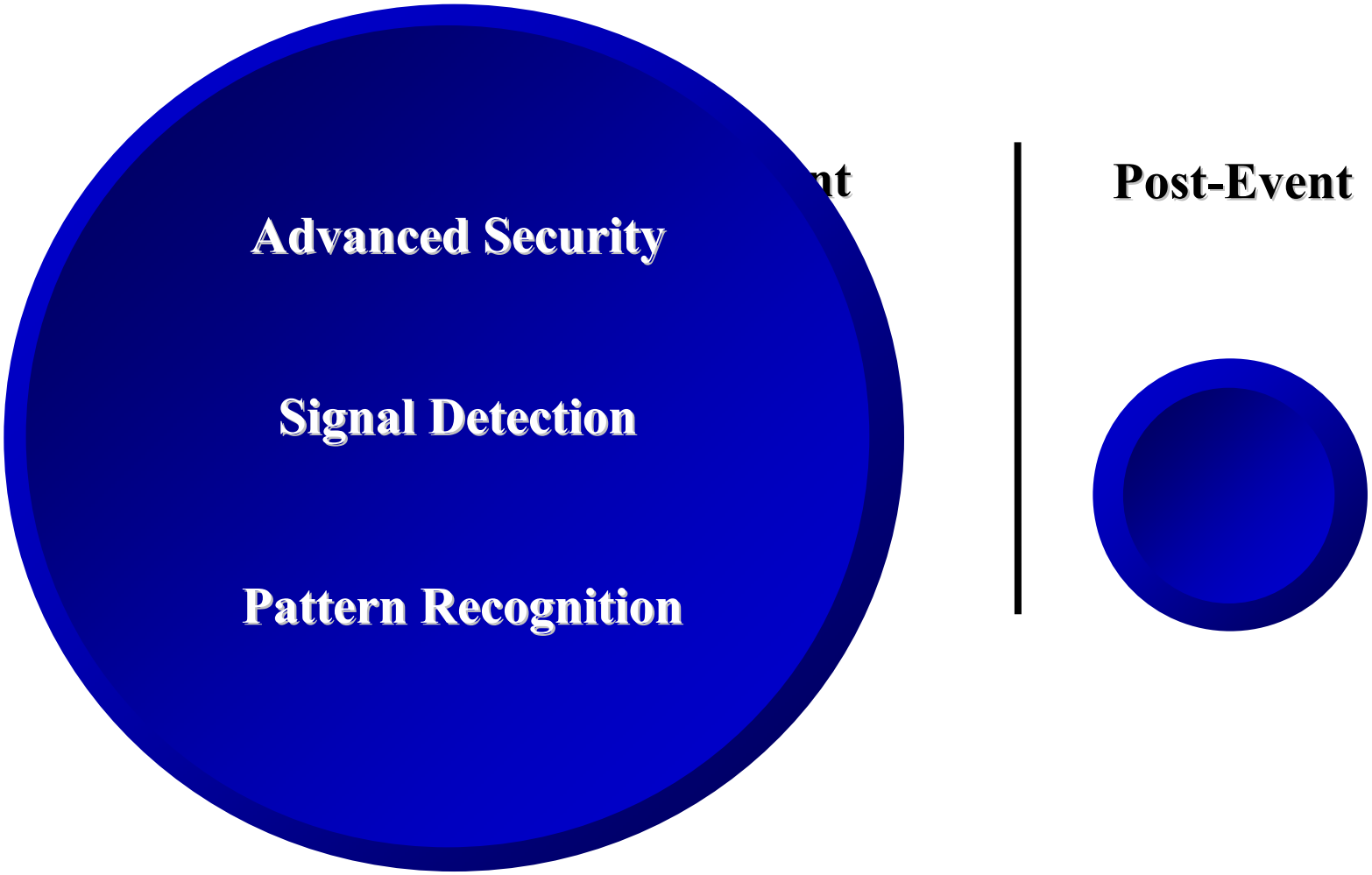


Adapted from Haddon's Matrix

Human Factors Contributions to Homeland Defense



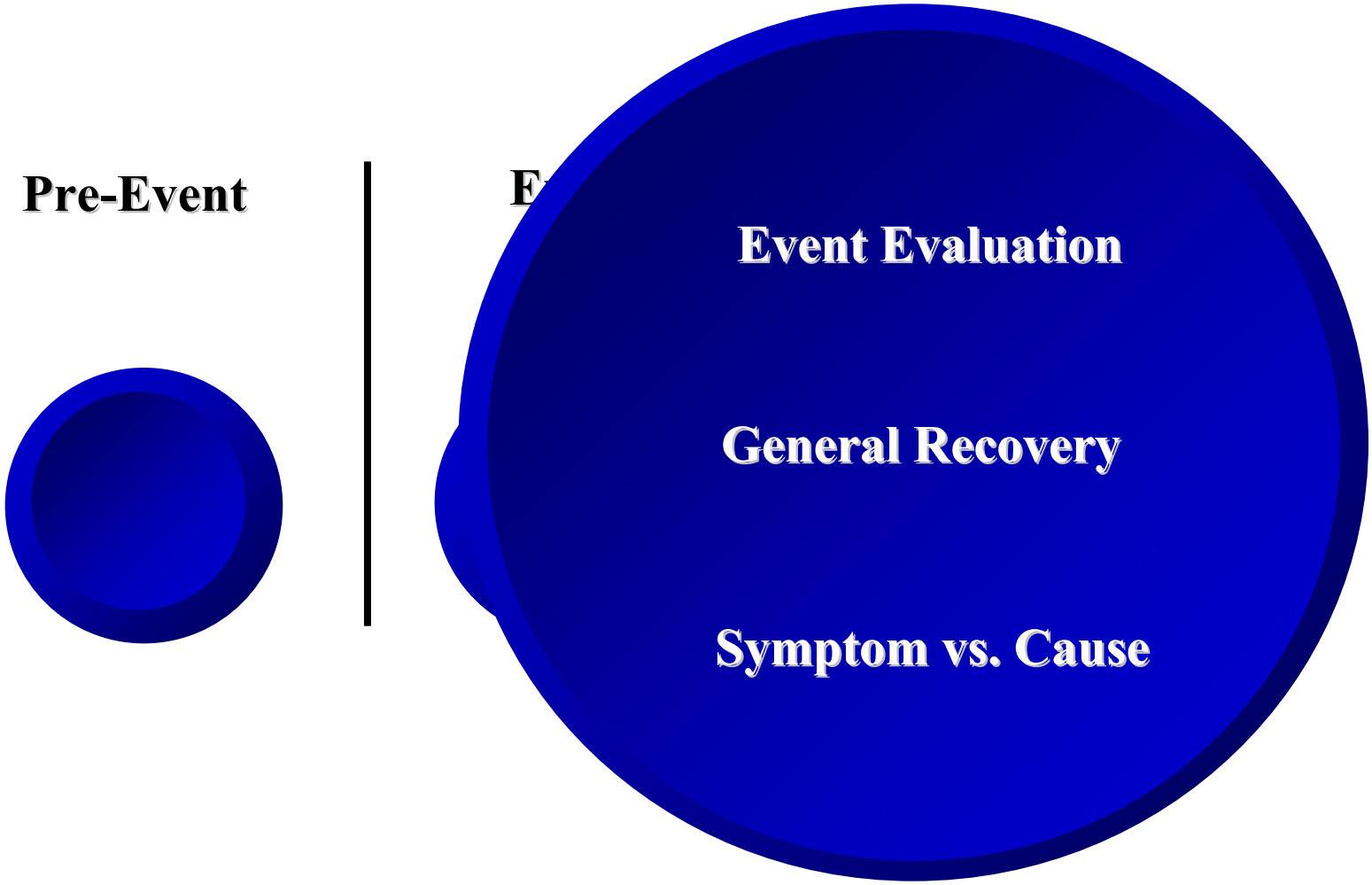
Human Factors Contributions to Homeland Defense



Human Factors Contributions to Homeland Defense

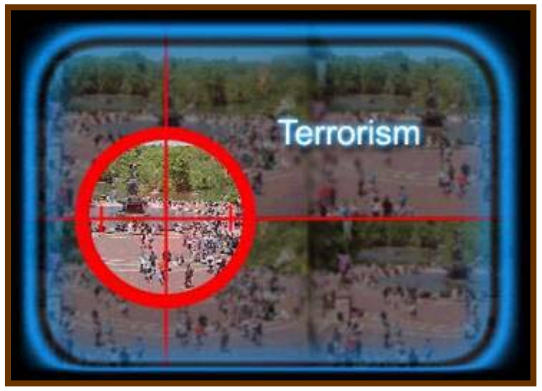


Human Factors Contributions to Homeland Defense



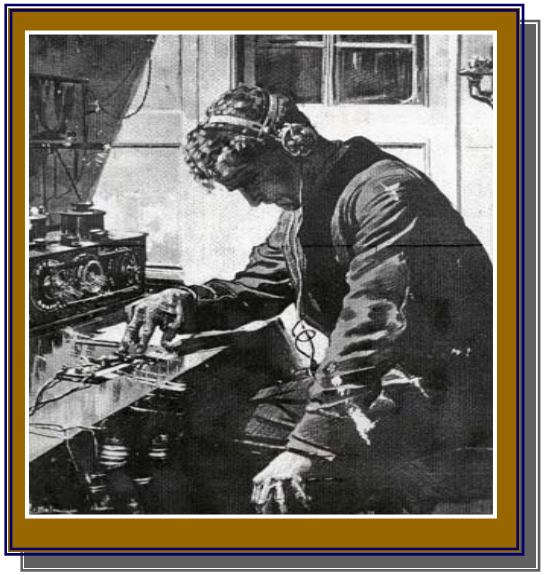
Human Factors Contributions to Homeland Defense

- Security Concerns
  - Vigilance/Sustained Attention
  - Stress
  - Fatigue and Workload
  - Training and Team Performance





The Problem of Vigilance/Sustained Attention



Vigilance is a Long-Standing Problem.

Sources of Performance Influence Include:

Event Rate

Signal Salience

Stress/Workload/Fatigue

Glare, Noise, Temperature, Vibration, TOD, Drug Effects etc

Memory Load

Successive vs. Simultaneous Comparisons

Feedback

Hit vs. Miss vs. FA KR/KP

Individual Differences

Introversion/Extraversion, Age, Sex, Expertise

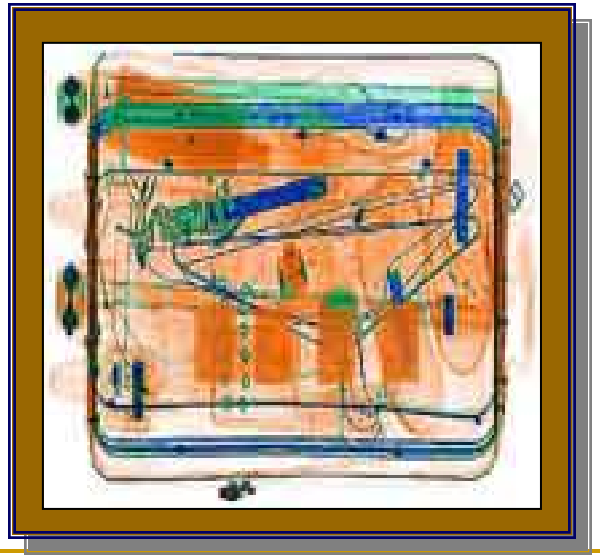
Warm, J.S. (Ed.). (1984). *Sustained attention in human performance.* New York: Wiley.

The Problem of Vigilance/Sustained Attention









Davies, D.R., & Parasuraman, R. (1982). *The psychology of vigilance*. Academic Press: London.

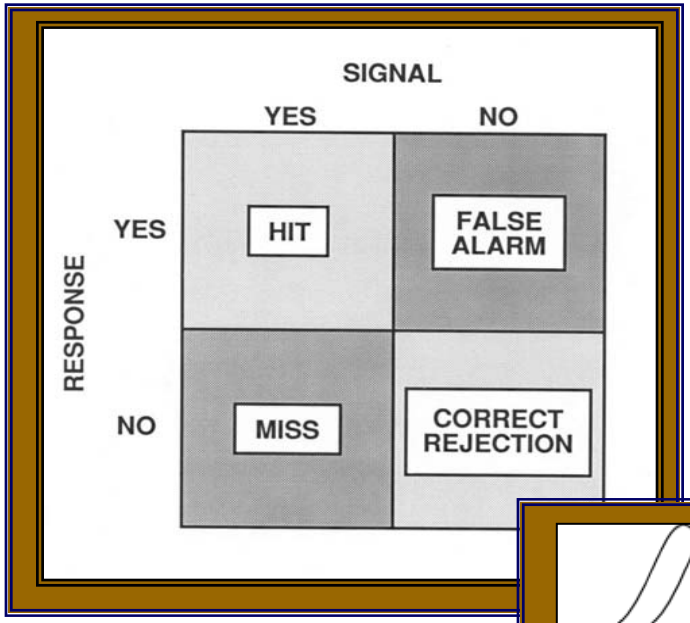
Body Image Analysis

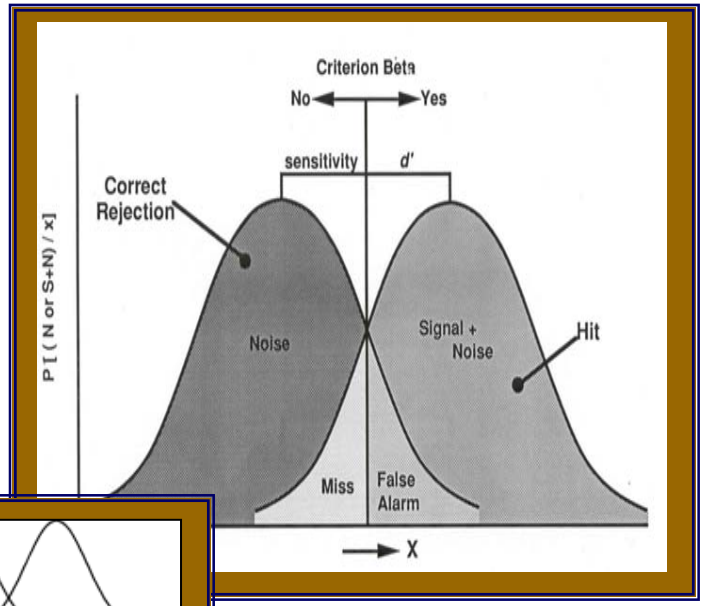


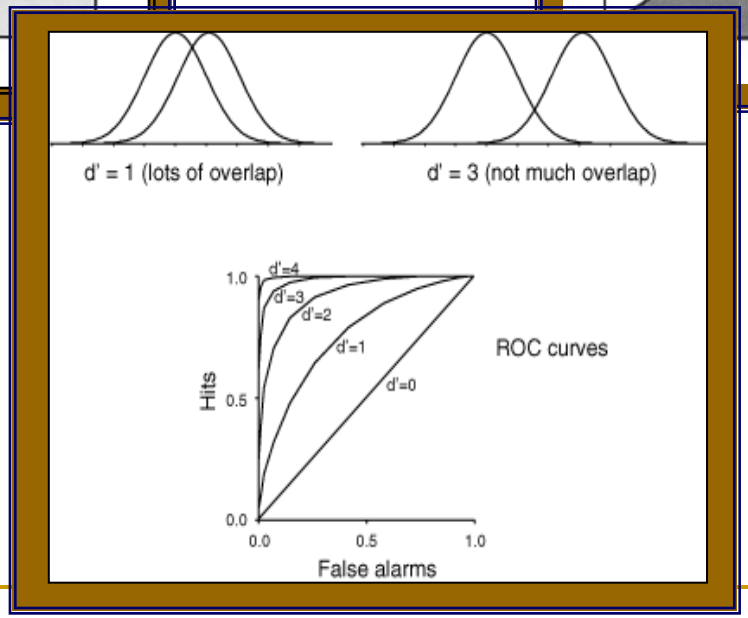




The Utility of Signal Detection







Swets, J.A. (1996). *Signal detection*. Erlbaum: Mahwah, NJ.

Human Factors Contributions to Homeland Security



- Military Response
  - Human Performance Modeling and Simulation
  - Performance Under Stress/High Workload



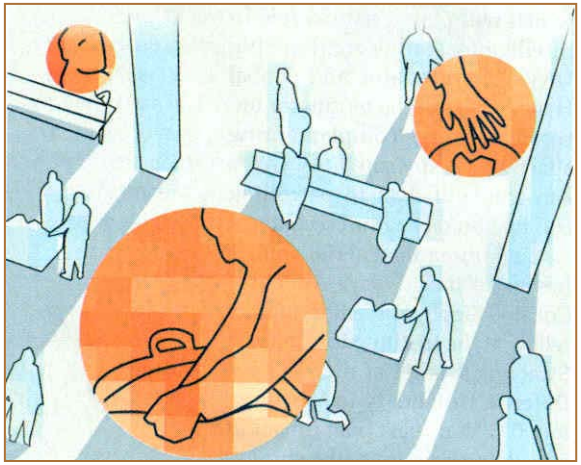
Human Factors Contributions to Countering Terrorism



- Technological Test and Development
  - Pattern Recognition
  - Template Matching Procedures
  - Iris scanning and/or face matching



Human Factors Contributions to Countering Terrorism



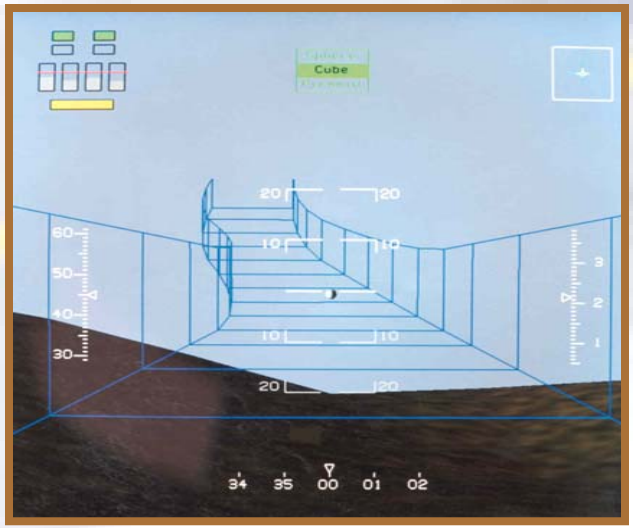
- Technological Test and Development
  - Behavior Recognition

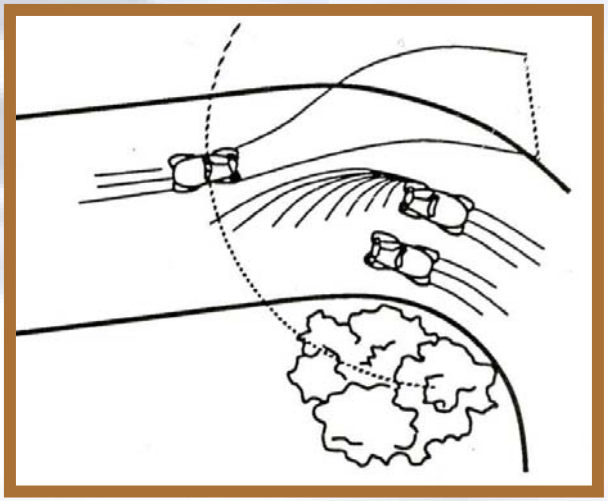
Human Factors Contributions to Homeland Security

‘Envelopes of Protection’ (Fields of Safe Travel)

‘Highways/No Go Areas in the Sky’

Cabin-to-Cockpit to Ground Video Link





Hancock, P.A. (2001). ***The tale of a two-face tiger. Ergonomics in Design, in press.***

Human Factors Interfaces to Search and Rescue Robots







Perceptual Issues

Orientation
- Scaling
- Displays
- Visual Workload
- **HUD Disorientation**

Movie Showing Robotic Exploration of Collapsed WTC Floors.



Control Issues

- Gain
- **Physical Ergonomics** (e.g., Gloves)
- Interaction with Semi-Automated System.
- etc

Mouloua, M., Gilson, R., & Hancock, P.A. (2001). Designing controls for future unmanned aerial vehicles. in press.

Photos and Movie Courtesy of the Center for Robot-Assisted Search and Rescue, University of South Florida.

Emergency Egress

- **Behavioral Insights into Evacuation Procedures (Courtesy of Jake Pauls).**
  - Movement of People. *The SFPE Handbook of Fire Protection Engineering Second Edition,* Society of Fire Protection Engineers and National Fire Protection Association, Section 3, Chapter 13, 1995, 3-263 - 3-285.
  - Building Evacuation: Research Methods and Case Studies. *Fires and Human Behaviour,* (Edited by D. Canter) Wiley , 1980, 227-250, (Coauthored with B. K. Jones).
  - Building Evacuation: Research Findings and Recommendations. *Fires and Human Behaviour,* (Edited by D. Canter) Wiley, 1980, 251-275.

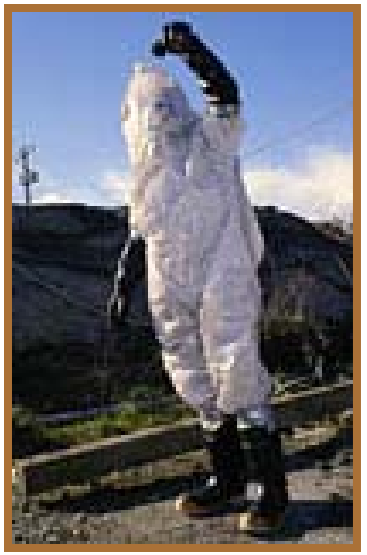
Pauls, J. (1994). Vertical Evacuation in Large Buildings: Missed Opportunities for Research. *Disaster Management*, 6 (3), 128-132.

Emergency Egress

- Non-ambulatory individuals
- Evacuation routes
  - Signage
  - Lighting
  - Crowd Control
- Rescue operations
  - Non-interfering movement streams
  - Situational Awareness



Human Factors Contributions to Homeland Defense



Physical Ergonomics

Combined Physical and Cognitive Demands

- Emergency Personnel
- Military Activities in an Asymmetric Warfare Scenario

Physical Activities in Confined Spaces





Human Factors Contributions to Homeland Defense



Multiple HF/E Problems of NBC and Emergency Equipped Personnel.

- **Physical Demand**
- **Heat Stress**
- **Interface Problems**
- **Restricted Vision**
- **etc**

Human Factors Solutions For Antecedent Conditions

Addressing The Poverty and Hopelessness that Generates Individuals who believe they have nothing to lose.

HF/E Solutions For Food/Water Supplies

HF/E Solutions to Infrastructure Re-Building







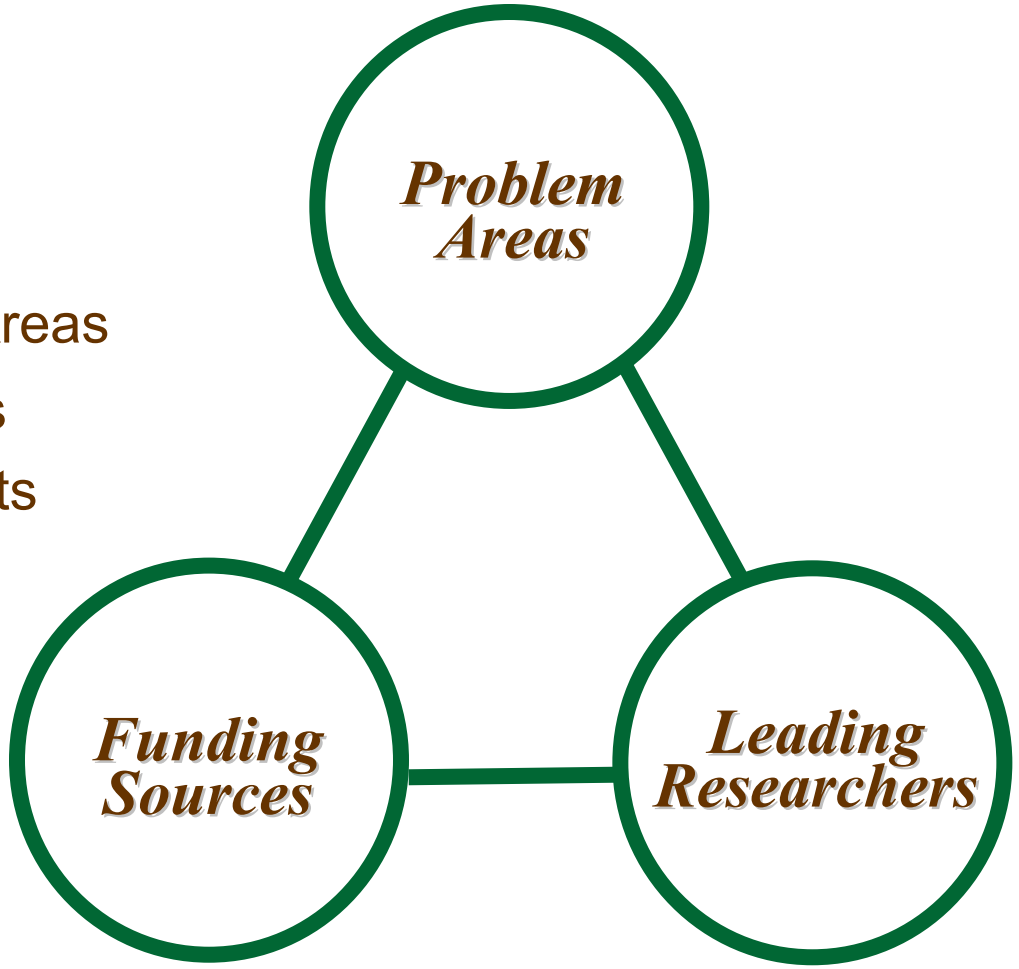
Nickerson, R. (1992). Looking ahead: Human Factors challenges in a changing world. Erlbaum: Hillsdale, NJ.

Moray, N. (1994). Ergonomics and the global problems of the 21st Century. Keynote Address, International Ergonomics Society, Toronto.

Human Factors Contributions to Homeland Defense

Action Items

- Identify Major Problem Areas
- Identify Funding Sources
- Identify Leading Scientists



Acknowledgements

This research was facilitated by to grants from the United States Army. The first was the Multiple University Research Initiative-Operator Performance under Stress (MURI-OPUS) Grant (#DAAD19-01-0621). The second was to the Advanced Decision Architecture Consortium (# DAAD 19-01-0009). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense or the U.S. Government. The authors would like to thank Dr. Sherry Tove and Dr. Mike Drillings for providing administration and technical direction for the first grant and to Dr. Mike Strub for the second grant.



Applications of Fuzzy Signal Detection Theory to Law Enforcement

J.L. Szalma, P.A. Hancock, J.L. Weaver, M. Mouloua, & S.C. Stafford

Abstract

The assessment of threat level and decisions regarding how to respond to that threat are critical facets of daily police work. While analysis of threat detection is amenable to signal detection theory, there are aspects of police work that are not captured by the model. In policing there are multiple levels of critical events, each of which requires a unique response. A police officer cannot afford to label subjects according to discrete categories of threats or non-threats, as these levels of threat lie along a continuum. In addition, policing agents are required to use levels of force that correspond to the level of threat detected, or in some cases elevate the level of force one category above that used by the subject. Hence, when responses by police officers to suspicious or threatening events are reduced to binary decisions (as in signal detection theory), critical information regarding the intent of the subject in those situations may be lost. This can have deleterious effects on police officers' lives and police-community relations. A new approach to signal detection theory that captures the continuous nature of the threat dimension is applied here to issues relevant for law enforcement. Initial empirical tests of the model that have implications for law enforcement are also discussed.

Introduction

Traditional Signal Detection Theory (SDT; Green & Swets, 1966) requires categorization of persons or events into one of two categories: signal (threat) versus non-signal (non-threat). However, such categorization limits analysis of complex stimuli and responses such as those in law enforcement. A recent advance in SDT promises to enhance measurement of performance in such complex real world domains. This development, Fuzzy Signal Detection Theory (FSDT; Hancock, Masalonis, & Parasuraman, 2000; Parasuraman, Masalonis, & Hancock, 2000), combines traditional SDT with Fuzzy Set Theory to extend signal detection analysis beyond the traditional crisp, categorical decision making model. FSDT allows for events to simultaneously be in more than one state category (e.g. signal and non-signal), so that the stimulus and response dimensions can be continuous rather than categorical. Consequently, FSDT can be used in applied settings such as law enforcement, where the degree of threat and the appropriate response to it are constantly changing. This work aims to specify the ways in which FSDT can be applied to the problems faced by those designing law enforcement systems and police training programs. Critical to this application is the identification of the mapping between the signal dimension and the state-of-the-world in a continuously changing environment in which the definition of a 'signal' changes over time and situations.

Fuzzy Stimulus Dimension (Threat)

The level of threat police officers face are often dynamic, varying both within and across work shifts. Police Officers are required to continuously adapt to these changes and adjust their responding according to the level of threat in their environment. The threat level can vary according to environmental conditions (e.g., day versus night) and the psychological characteristics (state and trait) of officers, civilians, and suspects.

Some Key Factors Affecting Level of Threat

- **Residential Districts**: Extreme High Income, High Income, Middle Income, Low Income, Extreme Low Income
- **Environment**: Day Patrol, Night Patrol, Extreme Weather Shifts (Mid-Summer Heat Advisory), (Winter Freeze Warnings), etc.
- **Elements of the Patrol Squad**: Number of officers available, Availability of Specialized Units (Canine, Traffic, Air Patrol), Availability of Back-up units, etc.
- **Political Climate**: March 3rd 1991 (Rodney King), Post September 11th. The General Public view of a Police Department following any incidents that are controversial or in the Media, etc.
- **Nature of the Crime or Offense**: Traffic Infraction, DUI, Assault, Battery, Aggravated Assault, Aggravated Battery, Rape, Homicide, etc.
- **Physical & Psychological State of Police Officers and Subjects**: The extent to which Officers and Subjects are physically capable of strenuous physical activity and the mental state (e.g., stress, workload) of these individuals.

University of Central Florida — Multi-Disciplinary University Research Initiative — MURI-OPUS — Operator Performance Under Stress — www.mit.ucf.edu

Correct Identification of the Stimulus Dimension (Threat)

Levels of Threat: (The Stimulus Dimension)

• Police Officers are presented with various levels of threat ranging from a non-threatening situation (a signal value of 0) to the maximum level of threat such as the death of an individual (a signal value of 1). Correctly assessing the degree of threat is a major component of Officer Safety and Survival training.

Levels of Force: (The Response Dimension)

• Levels of Force (Responding to a Stimulus) in Law Enforcement are determined by the threat level perceived by the officer to him/herself, citizens, subject, and/or property. Officers are trained to respond to situations by using levels of force appropriate to the nature of the crime, incident, or circumstances. The level of force can range from passive observing (response value of 0) to deadly force (response value of 1).

Stimulus (Threat) and Response Dimension (Action): Elements of Fuzzy Signal Detection

Traditional Set Theory: Set membership is mutually exclusive: elements either are or are not members of a set.

Fuzzy Set Theory: Events can simultaneously be and not be members of a set.

In Fuzzy Signal Detection Theory, either the stimulus (**s**) or response (**r**) dimensions (or both) are continuous. If both **s** and **r** are binary, FSDT reduces to the traditional (crisp) SDT model. A continuous **s** dimension varies from 0 (a non-signal) to 1 (a signal), with a signal value of .5 representing maximal uncertainty in the 'signalness' of the stimulus. Similarly, a continuous **r** dimension varies from 0 (a response of 'non-signal') to 1 (a response of 'signal'), with .5 representing maximal uncertainty in the response to the stimulus.

Multi-Disciplinary University Research Initiative — MURI-OPUS — Operator Performance Under Stress

1. Select mapping functions for signal & response dimensions

The mapping function can be discrete or continuous, and can be derived empirically, theoretically, or based upon legal or industry standards. Either **s** or **r** (or both) can be fuzzy. The measurement challenge is to map **s** to the actual states-of-the-world.

2. Assignment of Fuzzy Membership to SDT Categories using Mixed Implication Functions

►H = min (**s**,**r**) ►M = max (**s**-**r**, 0)
►FA = max (**r**-**s**, 0) ►CR = min (1-**s**, 1-**r**)

3. Computation of Fuzzy Hit and False Alarm Rate

► H= $\Sigma(Hi)/\Sigma(s_i)$ for i=1 to N ► M = $\Sigma(Mi)/\Sigma(s_i)$ for i =1 to N
► FA = $\Sigma(FA_i)/\Sigma(1\text{-}s_i)$ for i=1to N ►CR = $\Sigma(CR_i)/\Sigma(1\text{-}s_i)$ for i= 1 to N

4. Computation of Fuzzy Sensitivity and Response Bias Measures

Fuzzy sensitivity and response bias measures can be calculated in the traditional way, since the fuzziness of signal and response have been captured in the computations of H and FA.

e.g. $d' = Z_H - Z_{FA}$; $c = -.5(Z_H + Z_{FA})$.

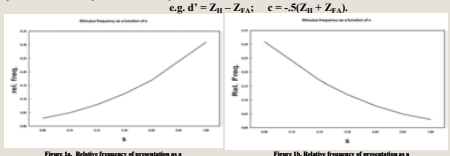


Figure 1a. Relative frequency of presentation as a function of 'signalness' for the 'high-signal' condition

Figure 1b. Relative frequency of presentation as a function of 'signalness' for the 'low-signal' condition



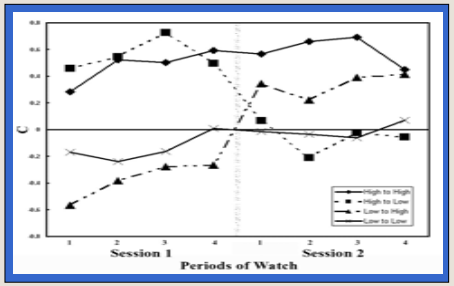


Figure 2. Response bias scores as a function of periods of watch for the four probability transition groups. Data for the two sessions are displayed separately in each panel. Note. High = greater frequency of 'signal-like' stimuli; Low= greater frequency of 'non-signal-like' stimuli (see Figure 1). Each period of watch was 5-min.

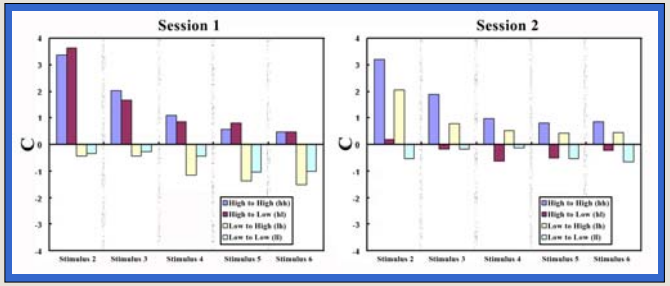


Figure 3. Response bias scores as a function of stimulus category for the four probability transition groups. Data for the two sessions are displayed separately in each panel. Note. High = greater frequency of 'signal-like' stimuli; Low= greater frequency of 'non-signal-like' stimuli (see Figure 1).

Responses to high versus low threat situations can have different effects on officer response bias, or the tendency for officers to respond with higher levels of force. Recent experimental work has demonstrated that the FSDT model is sensitive to manipulations of the relative frequency of stimuli that are high or low in threat or 'signalness' (Szalma, Hancock, Mouloua, & Stafford, 2003). Specifically, Szalma et. al. examined the impact of transitions on response bias by varying the frequency distributions for stimulus presentation in a vigilance task. The task required participants to monitor the degree to which two vertically aligned rectangles were similar in color. Observers engaged in two vigils. In the first 20-minute session they monitored a display in which either stimuli high in 'signalness' (i.e., the rectangles were different in color) were more frequent (see Figure 1a) or stimuli low in signal value were more frequent (see Figure 1b). In the second 20-minute session participants either experienced the condition to which they were assigned in the first session or they switched to the other distribution. Observers were not informed of the relative frequency of stimuli or the change between sessions. Details regarding the stimuli and task characteristics can be found in Szalma et al. (2003).

As can be seen in Figure 2, observers who switched from a high to a low frequency of 'signal-like' stimuli became more lenient in the second session than those who experienced the high condition in both sessions. There was a trend for those in the low-to-high transition group to become more conservative in the second session, but this trend was not significant (see Szalma et al., 2003).

When the data are analyzed separately for each stimulus, the effect of 'signal' frequency shifts on response bias in both sessions is more pronounced for stimuli that are more 'non-signal-like' (i.e. stimulus 2 and stimulus 3 in Figure 3). This represents an asymmetry in signal frequency effects for stimuli at different ends of the signal dimension. Further research is necessary to identify the factors that control this asymmetry.

Applications to Law Enforcement

The application of Fuzzy Signal Detection to Law Enforcement offers a unique view into the world of policing. Initial findings regarding response bias reveal that shifting between high and low frequency of signal-like events can influence responding in a FSDT task. When applied to law enforcement applications, these data suggest that police officers who are faced with situations in which most events are low in threat level would respond in a consistently lenient manner. That is, they would be more likely to respond at a higher level of force greater than that appropriate for the actual threat level. In contrast, police officers who are in situations where high threat levels are frequent are more likely to show restraint by responding at a lower level of force relative to the level of threat perceived. When police officers are shifted from high to low threat level environments the current data suggest that they will shift their responding to be more lenient in the low threat situation (i.e. they will respond with levels of force greater than that appropriate for the level of threat). In contrast, officers who respond to situations that are frequently low in threat level and then shift to responding to situations that are high in threat level do not seem to match officers responses that are always responding to high threat level situations. That is, officers who make this transition do not respond as conservatively as officers who are always responding to high threat level situations. If these data generalize to law enforcement domains, it is likely that training protocols would be necessary to shift these criteria to more appropriate levels. In particular, these methods will enhance training for community policing, where multiple levels of response are required of officers. In addition, such results would have implications for everyday law enforcement issues such as zone placement or job switching.

References

Green, D.M, & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hancock, P.A., Masalonis, A.J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science*, 1, 207-230.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, *42*, 636-659.

Szalma, J.L., Hancock, P.A., Mouloua, M., & Stafford, SC. (2003). Application of Fuzzy Signal Detection Theory to Vigilance: The Effect of Criterion Shifts. Paper submitted for the 47th Annual Meeting of the Human Factors and Ergonomics Society, Denver, CO.

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, P.A. Hancock, principal investigator. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.



University of Central Florida
Multi-Disciplinary University Research Initiative
MURI-OPUS
Operator Performance Under Stress
www.mit.ucf.edu

MIT²
INSTITUTE for SIMULATION & TRAINING

Application of Fuzzy Signal Detection Theory to Homeland Defense

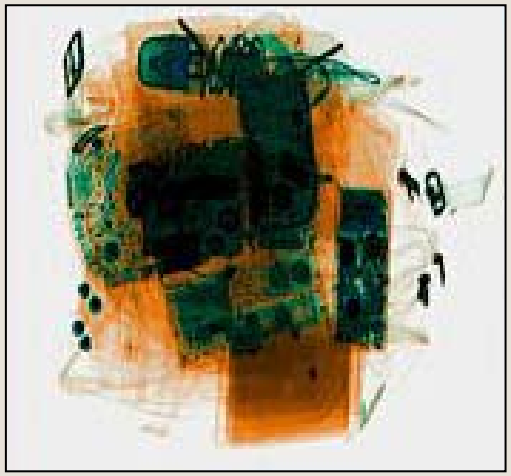
James L. Szalma

Peter A. Hancock

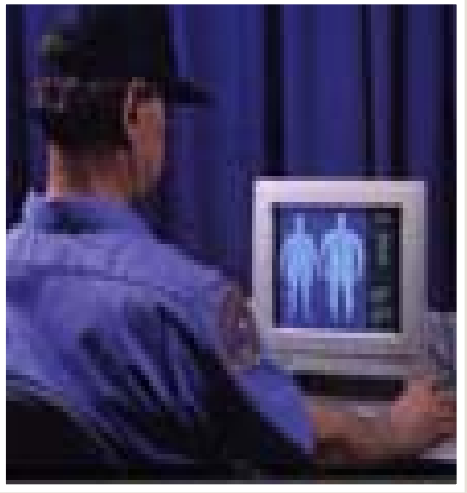
Mustapha Mouloua

University of Central Florida

Detection of Threat

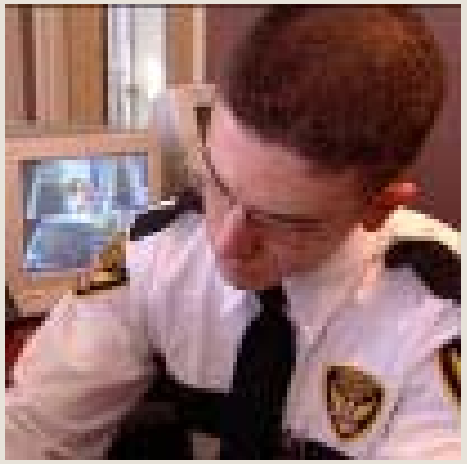












Human Performance under Stress

- Information Overload/Complexity
- Physical/Social Environment
- Uncertainty



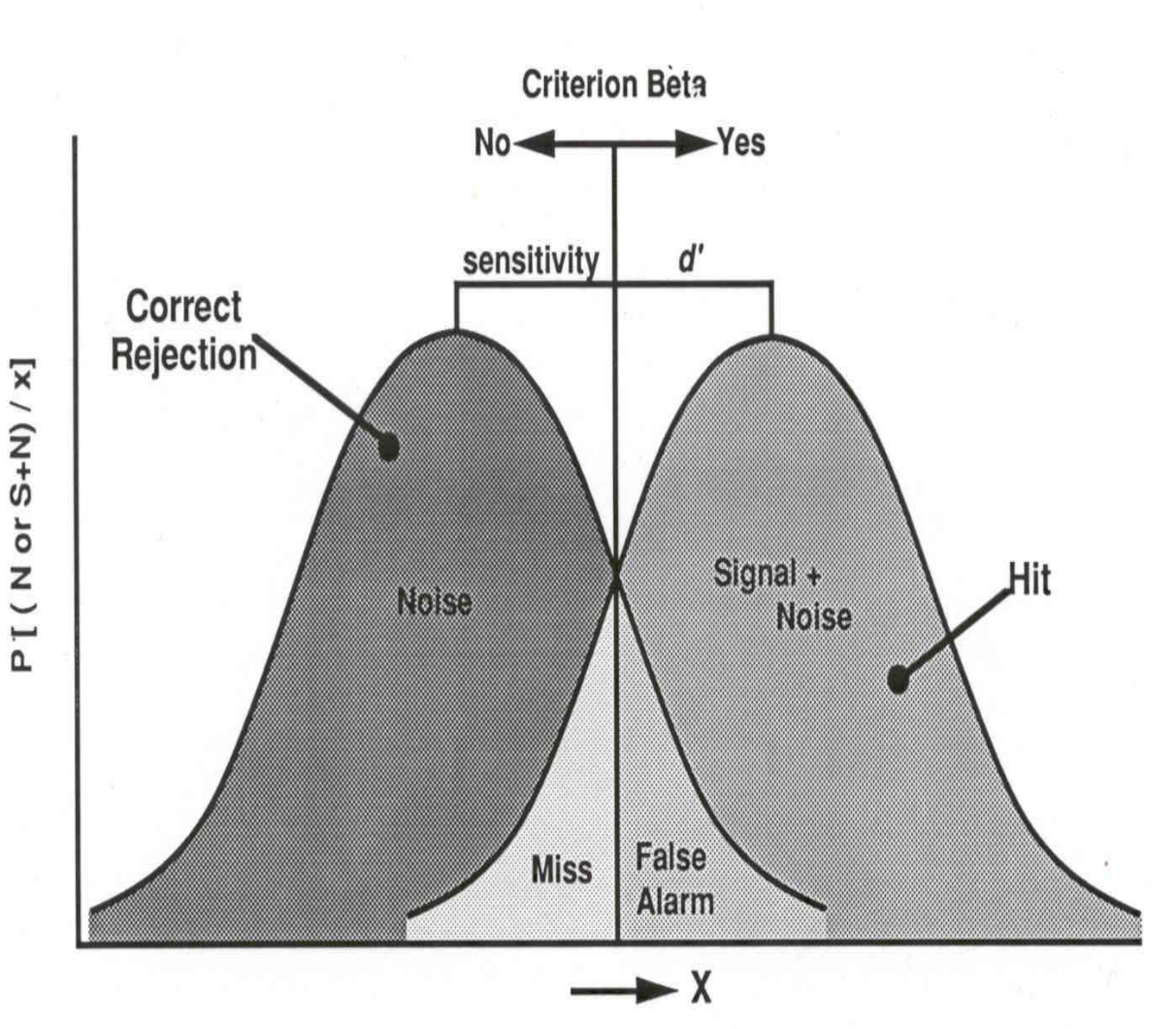




How Do We Assess Performance?

State of the World	Response: Yes	Response: No
Signal	Hit (H)	Miss (M)
Noise	False Alarm (FA)	Correct Rejection (CR)

(Crisp) Signal Detection Theory



Fuzzy Signal Detection Theory

Traditional Set Theory: $A \cap \bar{A} = 0$

Fuzzy Set Theory: $A \cap \bar{A} \neq 0$

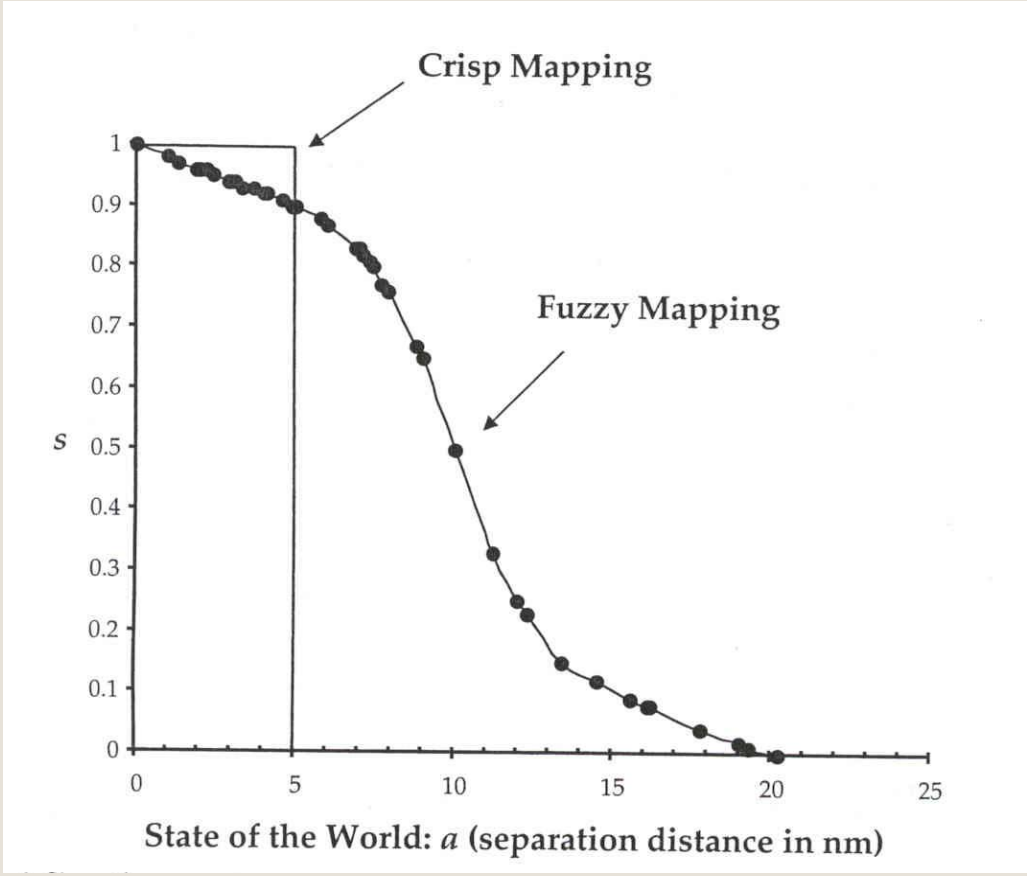
Degrees of membership possible

Elements of Fuzzy Signal Detection Theory

- Events can belong to the set “signal” (**s**) to a degree ranging from 0 to 1
- Events can belong to the set “response” (**r**) to a degree ranging from 0 to 1

Computation of FSDT Measures

- Select mapping functions for signal & response dimensions



Computation of FSDT Measures

- Assignment of degrees of membership to the four outcomes (H, M, FA, CR) using mixed implication functions.

$$H = \min(\mathbf{s}, \mathbf{r})$$
$$M = \max(\mathbf{s}-\mathbf{r}, 0)$$
$$FA = \max(\mathbf{r}-\mathbf{s}, 0)$$
$$CR = \min(1-\mathbf{s}, 1-\mathbf{r})$$

Parasuraman, Masalonis, & Hancock (2000)

- Note: Other mixed implication functions possible

Computation of FSDT Measures

- Compute fuzzy Hit, Miss, False Alarm, and Correct Rejection Rates (sum over the entire stimulus dimension)

- Compute detection theory measures of sensitivity and response bias

Note:

- One can analyze multiple categories of stimuli

- d' and c can be computed for a single trial or stimulus category

Continuous ***s*** with binary ***r***

In this case, the mixed implication functions reduce to:

If r=yes (1), then

Hit = **s**
Miss = 0
FA = 1-**s**
CR = 0

If r= no (0), then

Hit=0
Miss=**s**
FA=0
CR=1-**s**

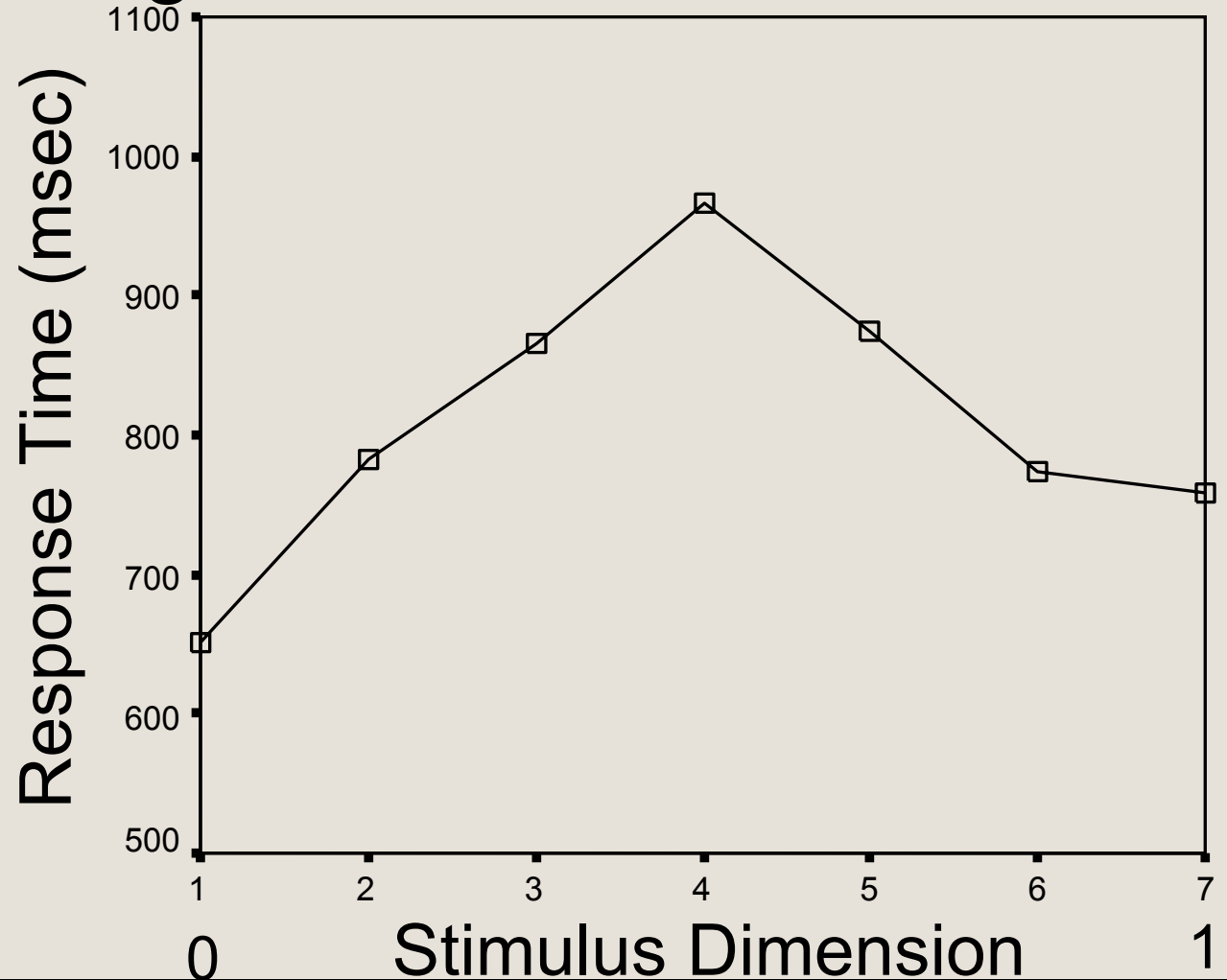
Current Empirical Work

Construct Fuzzy ROC Curves

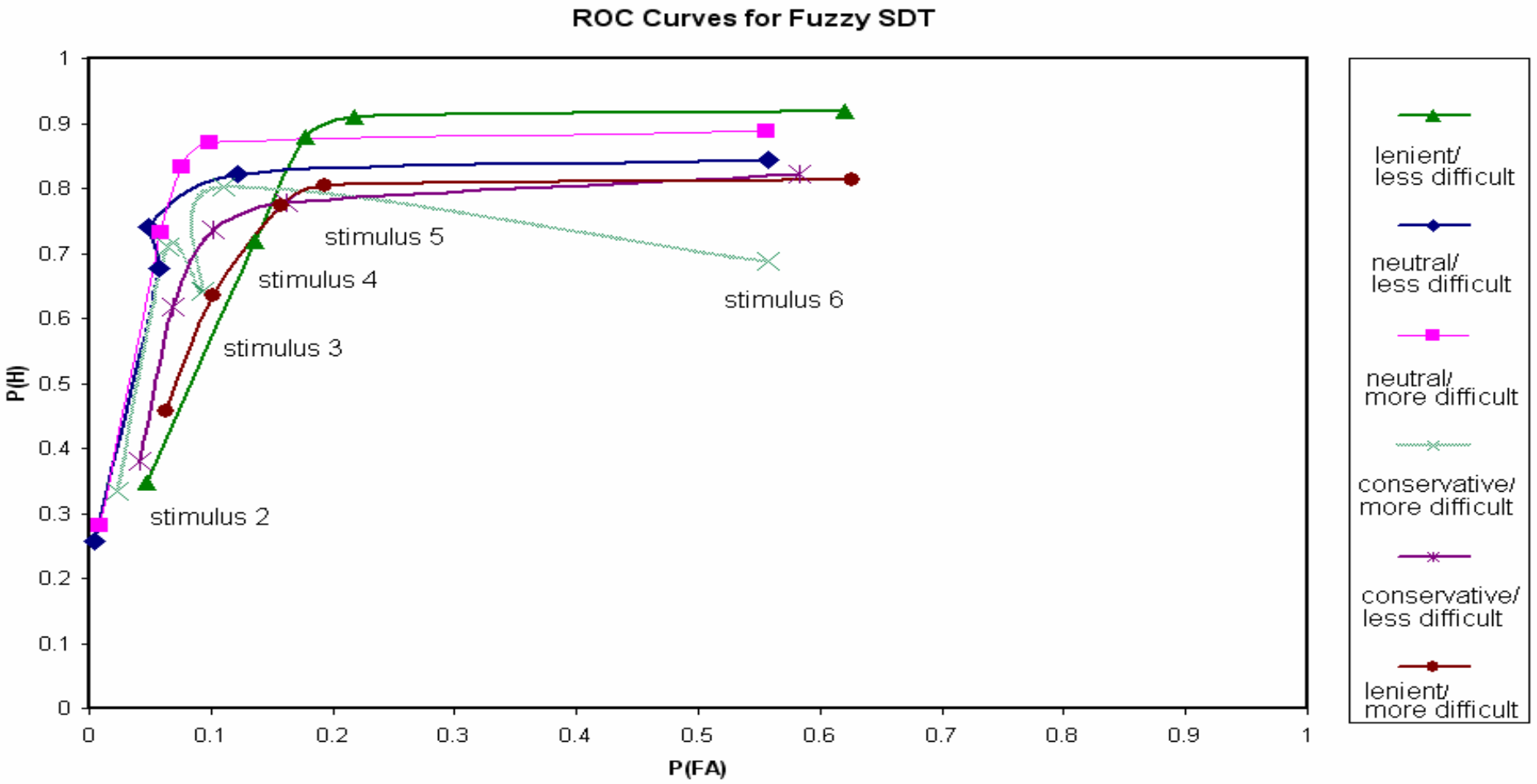
Compare to Crisp Model

FSDT is more sensitive to stimulus uncertainty

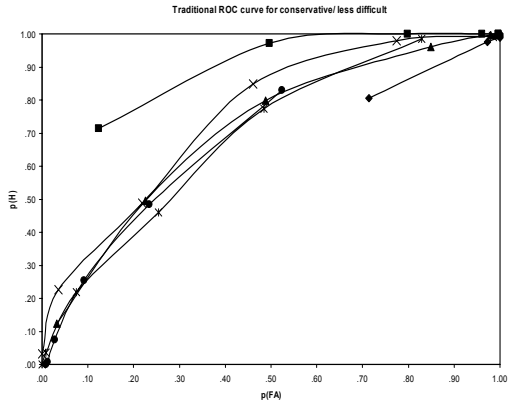
Response Time as a Function of Degree of Stimulus Uncertainty

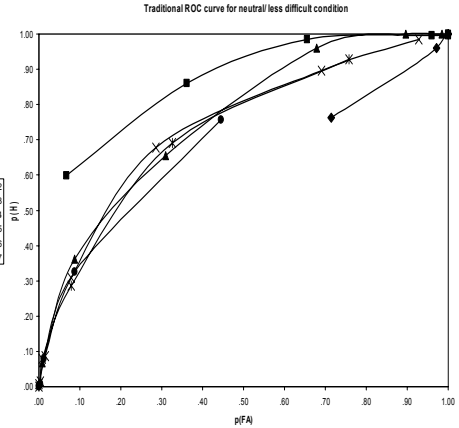


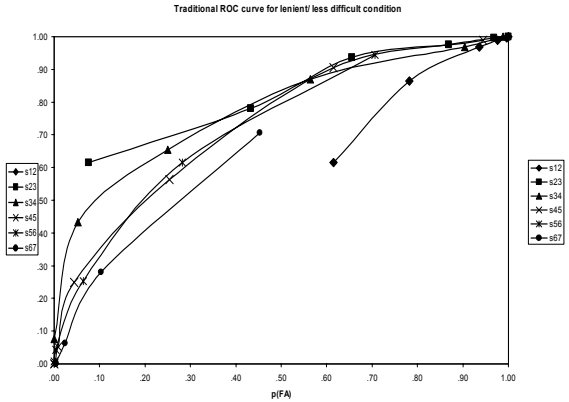
A Fuzzy ROC

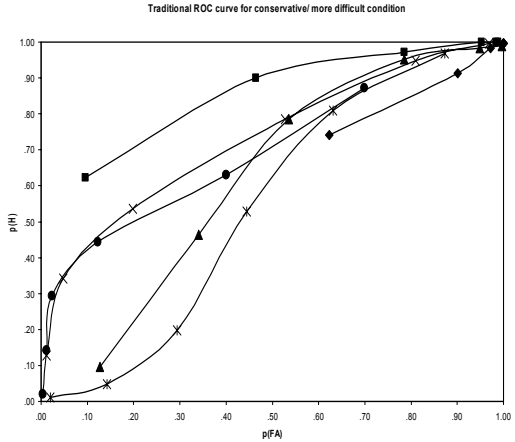

ROC Curves for Fuzzy SDT
P(H)
P(FA)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
stimulus 2
stimulus 3
stimulus 4
stimulus 5
stimulus 6
lenient/ less difficult
neutral/ less difficult
neutral/ more difficult
conservative/ more difficult
conservative/ less difficult
lenient/ more difficult

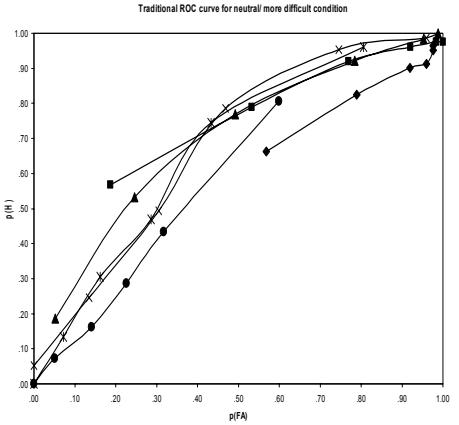
Multiple Stimuli/Response ROCs

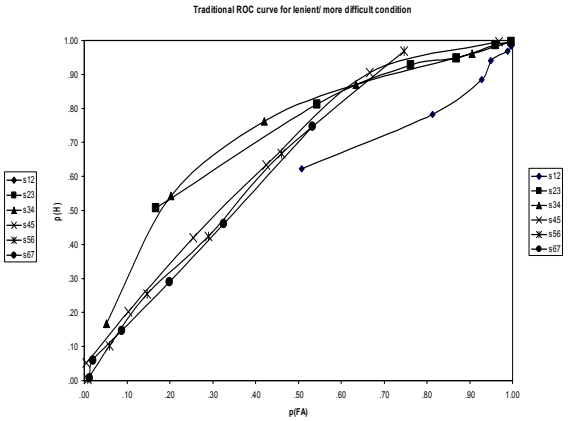












Current Empirical Work

Fuzzy Signal Detection and Vigilance

Shifts in 'signal' probability

Application of FSDT to Homeland Defense

- Vigilance (e.g. airport luggage screening: what constitutes a weapon?)
- Security Systems
- Transportation
- Terrorist identification/detection
- Threat assessment (e.g. based upon intelligence data)
  Why are we at level orange?
- Law Enforcement

Conclusions

- FSDT can improve the assessment of current and potential systems in response to terrorism.
- This method can be applied across several domains
- FSDT accounts for the uncertainty not only in the observer, but in the state of the world itself.
- Further empirical work is needed to evaluate the FSDT model.

ACKNOWLEDGEMENT

This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program , P.A. Hancock, principal investigator, administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the author and do not necessarily reflect official Army policy. The author wishes to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

Individual and Dual Task Performance Decrements as a Function of Stressor Exposure

Kristy A. Bradley, Jeanne L. Weaver, Adams Greenwood-Ericksen, Jessica A. Helmick,
Kelly A. Burke, and Peter A. Hancock
University of Central Florida



Abstract

An important aspect of soldier effectiveness is the ability to perform multiple tasks in the face of stress. It has likewise been found that performance on a secondary task degrades when multiple tasks are being performed simultaneously and that the presence of stress is associated with general decrements in performance. Consequently, it is important to understand how the presence of stress impacts complex task performance in the military domain. The purpose of the present study was to investigate the effect of noise on multiple task performance.

Introduction

Soldiers frequently are required to operate complex technology in stressful environments. It has been established that performance on a secondary task degrades when multiple tasks are performed simultaneously. Further, the presence of stress is associated with general decrements in performance (Hancock & Desmond, 2001; Matthews, Davies, Westerman, & Stammers, 2000). In order to alleviate these adverse effects, it is important to understand how the presence of stress impacts complex task performance. The present study is the first in a series that investigates the effects of various stressors on multiple task performance. We used the Multi-attribute Task Battery (MATB; Comstock & Arnegard, 1992) composed of skill, rule, and knowledge based activities (Rasmussen, 1983) to evaluate performance variation as a function of exposure to noise.

An integral component of Rasmussen's (1983) model is attention. Two current theories of stress incorporate the importance of the delegation of limited attentional resources in response to stressors, each in a slightly different manner. Lazarus and Folkman's (1984) transactional model of stress places emphasis on the interaction between the person and the demands within the environment. Hancock and Warm (1989) propose a dynamic model of stress in which a normative zone is surrounded by a comfort zone. It is within this area that a system or individual is believed to perform at a normal target value.

In order to investigate the effects of stress on multiple task performance, the Multi-attribute Task Battery (MATB; Comstock & Arnegard, 1992) was used. The MATB is a computer-based task designed to evaluate operator performance and workload. The MATB requires the simultaneous performance of monitoring, dynamic resource management, and tracking tasks. The simultaneous performance of multiple tasks is a central feature of the MATB and it is this feature that is consistent with most operational systems and thus makes the task useful for our purposes as a research platform. Each of the MATB tasks is postulated to correspond with a specific performance category in relation to Rasmussen's human performance typology. The MATB tracking task requires the manual operation of a joystick in order to keep a target in the center of a tracking window. This tracking task is performed based on a relatively automated pattern of behaviors that require little if any conscious attention and the task is therefore classified as a *skill-based* behavior. The monitoring task requires the resetting of a gauge once it has been detected as operating out of a specified range. Specific actions are taken in response to each gauge in accord with an *a priori* determined set of rules. As such, this task is classified as a *rule-based* behavior. The resource management task requires the activation of "pumps" in order to fill two tanks and also requires the maintenance of certain levels within them. This dynamic, *knowledge-based* task has no predetermined rules for specific operation and therefore requires conscious processing due to the ongoing problem solving nature of the task.

Hypotheses

Hypotheses regarding the current study are developed in congruence with the aforementioned notion that one must draw from a limited pool of resources in order to respond to conditions of stress. It is hypothesized that those tasks requiring more attentional resources will be the earliest to be shed, or degrade in terms of performance, in response to a stressor, whereas tasks more routinized will remain relatively undisturbed due to the lack of necessity for attentional resources. Specifically, it is hypothesized that the resource management task classified as knowledge-based will be shed first in response to noise, followed by the monitoring task classified as rule-based, and the skill-based tracking task will be the last to be shed.

Method

Participants

Forty-one undergraduate students (11 men and 30 women, mean age = 20.10) who volunteered received extra credit to participate in the study.

Apparatus

The current study used a modified version of the original MATB to investigate the effects of noise on multi-task performance. The task labels on the screen of the MATB were modified to resemble duties associated with military tank operation. An aversive artillery noise delivered through headphones at 85dB in random intervals was presented as the stressor. Performance data was collected along with data from the NASA TLX (Hart & Staveland, 1998), an instrument that measures perceived workload. Data was also collected from a demographics questionnaire, the Dundee Stress State Questionnaire (DSSQ; Matthews, Joyner, Gilliland, Campbell, Huggins, & Falconer, 1999), the Cognitive Failures Questionnaire (CFQ; Broadbent, Cooper, FitzGerald, & Parks, 1982), and the Immersive Tendencies Questionnaire (ITQ; Witmer & Singer, 1998). The questionnaires were administered via Inquisit (Millisecond, 2000), psychological data collection software for Windows, on a Dell Dimension 8200 desktop computer.

Procedure

Participants were required to fill out the demographics questionnaire, the ITQ, the CFQ, and the DSSQ in Inquisit upon arrival. Following the questionnaires, each task of the MATB was presented individually and the participants practiced each task for five minutes. The order of individual task presentation was randomized across participants to avoid order effects. Upon completion of each task, feedback regarding performance was provided based on normative data generated from a previous pilot study. Participants also received feedback following a subsequent 10-minute practice session that combined all MATB tasks. Participants filled out the NASA TLX after the practice session with the assistance of the experimenter for practice purposes only.

Four observations of the combined MATB task performance were made with the occurrence of the stressor trial counterbalanced across participants to occur in either the second or third trial. Headphones were placed on the participant during each trial. During the noise stressor trial, the noise began 15 seconds after the start of the task. After each trial, participants filled out the DSSQ, NASA TLX, and a manipulation check question. The presentation order of the questionnaires was also counterbalanced across participants.

Results

As a manipulation check, a one-way within subjects analysis of variance was conducted to determine whether perceived stress was higher in the noise trial. A significant main effect of trial was observed, $F(3, 105) = 3.332$, $p = .022$ ($\eta 2 = .087$). Subsequent analyses indicated that perceived stress ratings in the noise condition ($M = 56.19$) were significantly greater than those in the control observation ($M = 47.22$), the first observation trial, ($M = 47.78$), and the final observation trial 4 ($M = 42.08$).

A mixed-model 2 (condition) x 4 (trial) analysis of variance was used to determine whether performance on each MATB task declined as a function of stressor exposure across trials. Order effects were tested by including condition as a between subjects variable and trial as a within subjects variable. A significant main effect of trial was found for tracking root mean square error, $F(2.341, 86.617) = 5.062$, $p = .006$ ($\eta 2 = .120$). Subsequent analyses indicated that tracking root mean square error for observation trial 1 ($M = 101.34$) was significantly greater than in both the stressor observation trial ($M = 95.33$) and the final observation trial 4 ($M = 92.94$). Root mean square error in the control observation trial ($M = 101.06$) was significantly greater than that of both the stressor observation trial ($M = 95.33$) and the final observation trial 4 ($M = 92.94$). No significant main effect of condition was observed , $F(1, 37) = 1.181$, $p = .284$ ($\eta 2 = .031$).

No significant main effects of trial were observed for: monitoring correct resets, $F(3, 105) = 1.557$, $p = .204$ ($\eta 2 = .043$), monitoring incorrect resets, $F(1.714, 65.113) = 1.403$, $p = .252$ ($\eta 2 = .036$), monitoring reset time, $F(3, 99) = 1.613$, $p = .191$ ($\eta 2 = .047$), resource management tank 1 deviation, $F(3, 102) = 1.920$, $p = .131$ ($\eta 2 = .053$), or resource management tank 2 deviation, $F(3,102) = 2.567$, $p = .059$ ($\eta 2 = .070$). No significant main effects of condition were observed indicating that there were no order effects.

***This research was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ARO under grant DAAD19-01-1-0621.**

Results Continued

A one-way within subjects analysis of variance was conducted to determine whether irritation levels increased across trial. A significant main effect of trial was found, $F(3.342, 133.680) = 7.901$, $p < .0005$ ($\eta 2 = .165$). Subsequent analyses revealed that irritation levels upon arrival ($M = 3.39$) were significantly lower than those after the first observation trial 1 ($M = 2.83$), the control observation trial ($M = 2.58$), the stressor observation trial ($M = 2.49$), and the final observation trial 4 ($M = 2.90$).

Pearson's r correlation coefficients were calculated among the ITQ, the CFQ, and each of the 5 performance measures across 4 trials. The ITQ correlated positively with the CFQ, r (N = 27) = .460, $p = .016$ (two-tailed) and the trial 1 observation of monitoring correct reset percentage, r (N = 27) = .432, $p = .024$ (two-tailed). The ITQ correlated negatively with stressor observation trial resource management tank 2 deviation, r (N = 27) = -.447, $p = .020$ (two-tailed), control observation trial resource management tank 2 deviation, r (N = 27) = -.392, $p = .043$ (two-tailed), trial 4 observation resource management tank 1 deviation, r (N = 27) = -.508, $p = .007$ (two-tailed), and trial 4 observation resource management tank 2 deviation, r (N = 27) = -.525, $p = .005$ (two-tailed). Correlations among the five different performance measures within each trial were also calculated. Results indicated that none of these correlations were significant. That is, there was no significant relationship among any of the performance measures within trial.

Discussion

In general, the performance findings did not support our original degradation hypotheses. Although the manipulation check regarding perceived stress showed that the experimental condition differed significantly from the other trials, there was only one significant performance finding. One possible explanation for this finding was that the task itself was so stressful that the addition of the noise failed to result in any meaningful difference. In order to test this possibility, the question regarding subjective irritation from the DSSQ was compared across the four task performance conditions with the response given by subjects upon their arrival. Interestingly, this comparison indicated that subjects were significantly less irritated upon arrival in comparison to all task performance trials. Thus, perhaps the task itself was consistently stressful enough in a performance context that performance was only minimally influenced.

The one significant performance finding related to tracking performance. However, tracking performance was *improved* by noise exposure. Although this was not an *a priori* hypothesized finding within the context of the current study, the findings of stress researchers such as Poulton (1978) provide insight regarding this result. It is likely that participants were innervated by the noise in this condition because they had been performing the task for at least 45 minutes. Thus, the noise countered subjective fatigue possibly being experienced by the participants at this time. Further analyses of our remaining data will likely shed further light on this finding. It is hoped that the findings of this research program will ultimately contribute to our understanding of the effects of various stressors on different types of tasks in the military domain. The current study serves to illustrate that stressor effects are often not intuitive and in fact often counter to what is hypothesized. Most important is to consider the utility of the MATB as truly illustrative of the components of Rasmussen's typology. Future efforts will be devoted to this and other issues (e.g., choice and manipulation of stressors) related to the continuation of this research program.

References

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21*, 1-16.

Comstock, J. L., & Arnegard, R. J. (1992). *The Multi-attribute Task Battery for human operator workload and strategic behavior research.* Technical Report 104174. Hampton, VA: NASA Langley Research Center.

Hancock, P. A., & Desmond, P. A. (2001). *Stress, workload, and fatigue.* Mahwah, NJ: Lawrence Erlbaum.

Hancock, P. A., & Warm. J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519-537.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North-Holland.

Millisecond Software (2002). Inquisit 1.XX [Computer software]. Seattle, WA: Millisecond Software. Lazarus, R. S., & Folkman, S. (1984). *Stress appraisal and coping.* New York: Springer.

Matthews, G., Davies, D. R., Westerman, S. J., & Stammers, R. B. (2000). *Human performance: cognition, stress, and individual differences.* Philadelphia: Taylor and Francis.

Matthews, G. Joyner, L., Gilliland, K., Campbell, S. Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (vol 7, pp. 335-350).

Poulton, E. C. (1978) A new look at the effects of noise: A rejoinder. *Psychological Bulletin, 85*, 1068-1079.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics, SMC 13*, 257-266.

Witmer, B. G., Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225-240.

Individual and Dual Task Performance Decrements as a Function of Stressor Exposure

Kristy A. Bradley, Jeanne L. Weaver, Adams Greenwood-Ericksen, Jessica A. Helmick,
Kelly A. Burke and Peter A. Hancock
University of Central Florida

An important aspect of soldier effectiveness is the ability to perform multiple tasks successfully, often times under stressful conditions. However, it has been found that performance on a secondary task degrades when multiple tasks are being performed simultaneously and that the presence of stress has been shown to be associated with decrements in performance (Matthews, Davies, Westerman, & Stammers, 2000). Consequently, it is important to understand how the presence of stress impacts complex task performance. The purpose of the current study is to investigate the effects of various stressors on multiple task performance. Specifically, the tasks of the Multi-attribute Task Battery (MATB; Comstock & Arnegard, 1992) classified as skill, rule, or knowledge based (Rasmussen, 1983) are being investigated in terms of the order in which they will shed, or degrade in terms of performance, as a function of exposure to various stressors. The MATB, modified to reflect tasks required of a tank operator, is a computer-based task designed to evaluate operator performance and workload. The MATB requires the simultaneous performance of monitoring, dynamic resource management, and tracking tasks and each of the MATB tasks can be considered to correspond with a specific performance category in relation to Rasmussen's human performance typology. The development of hypotheses regarding the current study is done so incongruence with the aforementioned notion that one must draw from a limited pool of resources in order to respond to conditions of stress. Specifically, we hypothesize that those tasks requiring more attentional resources will be shed, or degrade in terms of performance, first in response to a stressor, whereas tasks more routinized

will remain relatively undisturbed due to the lack of necessity for attentional resources. Specifically, it is hypothesized that the resource management task classified as knowledge-based will be shed first in response to noise, followed by the monitoring task classified as rule-based, and the skill-based tracking task will be the last to be shed. Data collected thus far indicates that the percentages of correct resets on the monitoring task were significantly higher in the individual training session (M = 100, SD = .00) than in the practice session in which all tasks were performed simultaneously (M = 71.67, SD =16.71), $t(7) = 4.795$, $p = .001$ (one-tailed). Monitoring task mean reset times were significantly higher in the practice session (M = 3.38, SD = .69) than in the individual training session (M = 2.67, SD = .78), $t(9) = -2.168$, $p = .029$ (one-tailed). These results support prior research indicating that time sharing task performance is typically inferior to individual task performance. Although mean differences between monitoring incorrect resets and tracking root mean square values during the individual training session and practice session approached significance, resource management tank target values and tank and tank activations failed to reach significance. Full results will be presented, including the effect of stress with regard to individual and dual task performance. Finally, the implications of these findings for military task performance under stress will be discussed.

Adaptive experts and automaticity: Explanations of failing under pressure

P. Ward, J.L. Szalma, and P.A. Hancock
University of Central Florida

The Hancock and Warm model (1989) predicts that under conditions of low-threat, a range of individual-specific problem solutions exist which result in non-stereotypical behavior. However, under stress, the number of available solution paths diminishes and behavior becomes more stereotypical as individuals narrow their attention to find the most appropriate solution. In more recent explications of this work, a number of alternatives to the question of how skill level influences attentional narrowing have been explored. Initial hypotheses speculated that tasks or processes that are reliant on declarative knowledge were likely to degrade under stress before rule-based processes. Those skills that were largely automated were predicted to be relatively resilient to stress (cf. Rasmussen, 1982). In contrast, the conscious control (Masters, 1992) and self-focus (Baumeister, 1984) hypotheses suggest that under stressful conditions, individuals attempt to explicitly monitor previously automated processes which may cause a shift towards controlled processing, inhibit automation and consequently, degrade performance. Lewis and Linder (1997) predicted that such a mechanism was a better explanation of the effects of stress upon performance than previous distraction hypotheses given that diversion away from explicit control processes under stress via a secondary distracter task ameliorated performance degradation. Recent examination of these predictions suggests that these effects may be dependent upon an individual's skill level or stage of learning. That is, distraction may degrade performance in low-skilled individuals in the early stage of learning where performance is typically cognitively-driven. Whereas, in the latter stages of learning explicit control of automated processes is likely to be explain failure under pressure in high skilled individuals (Beilock & Carr, 2001; Beilock et al., 2002). These predictions, however, are based on the assumption that expertise is simply a transition towards automaticity (Anderson, 1983; Fitts & Posner, 1967; Shiffrin & Schnieder, 1977). Whilst this assumption may hold for routine expertise, in reality experts in most real-world tasks are more appropriately described as 'adaptive' experts (Hatano & Ignaki, 1986). Adaptive experts acquire domain-specific skills that allow them to engage in explicit monitoring and control strategies which facilitate planning, evaluation, prediction and anticipation future events (see Ericsson & Kintsch, 1995; Ericsson & Lehman, 1996). Moreover, the regulatory control activities which mediate stress effects (e.g., Hockey, 1986) and the individual differences in predisposition toward threat (e.g., Eysenck, 1992) are likely to be integral to acquired expert representations and skills. The implications of mitigating the effects of stress through the acquisition of expertise are discussed.

Abstract presented at the ARO Bioscience Workshop, "Seeing Both the Scene and Unseen-Human Sight and Microscopic Sites," in Cashiers, NC April 2003.

References

Anderson, J. R. (1983). *The architecture of cognition Cambridge*, MA: Harvard University Press.

Baumeister, R. F. (1984). Choking under pressure: Self- consciousness and paradoxical effects of incentive on skillful performance. *Journal of Personality and Social Psychology*, *46*, 610-620.

Beilock, S. L. & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General, 130,* 701-725.

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied, 8,* 6-16.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211-245.

Ericsson, K.A., & Lehmann, A.C. (1966). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47,* 273-305.

Eysenck, M.W. (1992) Anxiety: The Cognitive Perspective. Hove: Lawrence Erlbaum Associates.

Fitts, P.M., & Posner, M.I. (1967). *Learning and skilled performance in human performance.* Belmont CA: Brock-Cole.

Hancock, P.A. & Warm, J. S. (1989). A dynamic model of stress and sustained attention. Human Factors, 31, 519-537.

Hatano, G, & Ignaki, K. (1986). Two courses of expertise. In *Child Development and Education in Japan*, H. Stevenson, H. Azuma, and K. Hakuta (Eds.). New York: W.H. Freeman.

Hockey, G.R.J. (1986). A state control model of adaptation and individual differences in stress management. In: G.R.J. Hockey, A.W.K. Gaillard & M.G.H. Coles (Eds.), *Energetics and human information processing*. (pp. 285-298). Dordrecht: Martinus Nijhoff.

Lewis, B. P. & Linder, D. E. (1997). Thinking about choking? Attentional processes and paradoxical performance. Personality and Social Psychology Bulletin, 23, 937-944.

Masters, R. S. W. (1992). Knowledge, knerves, and know-how. *British Journal of Psychology, 83*, 343-358.

Rasmussen, J. (1983). Skills, Rules, and Knowledge: Signals, signs, symbols and other distinctions in human performance models. IEEE Transactions on Systems: Man and cybernetics, 13.

Schneider, W. & R. M. Shiffrin. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review*, *84*, 1-66.

Abstract presented at the ARO Bioscience Workshop, "Seeing Both the Scene and Unseen-Human Sight and Microscopic Sites," in Cashiers, NC April 2003.

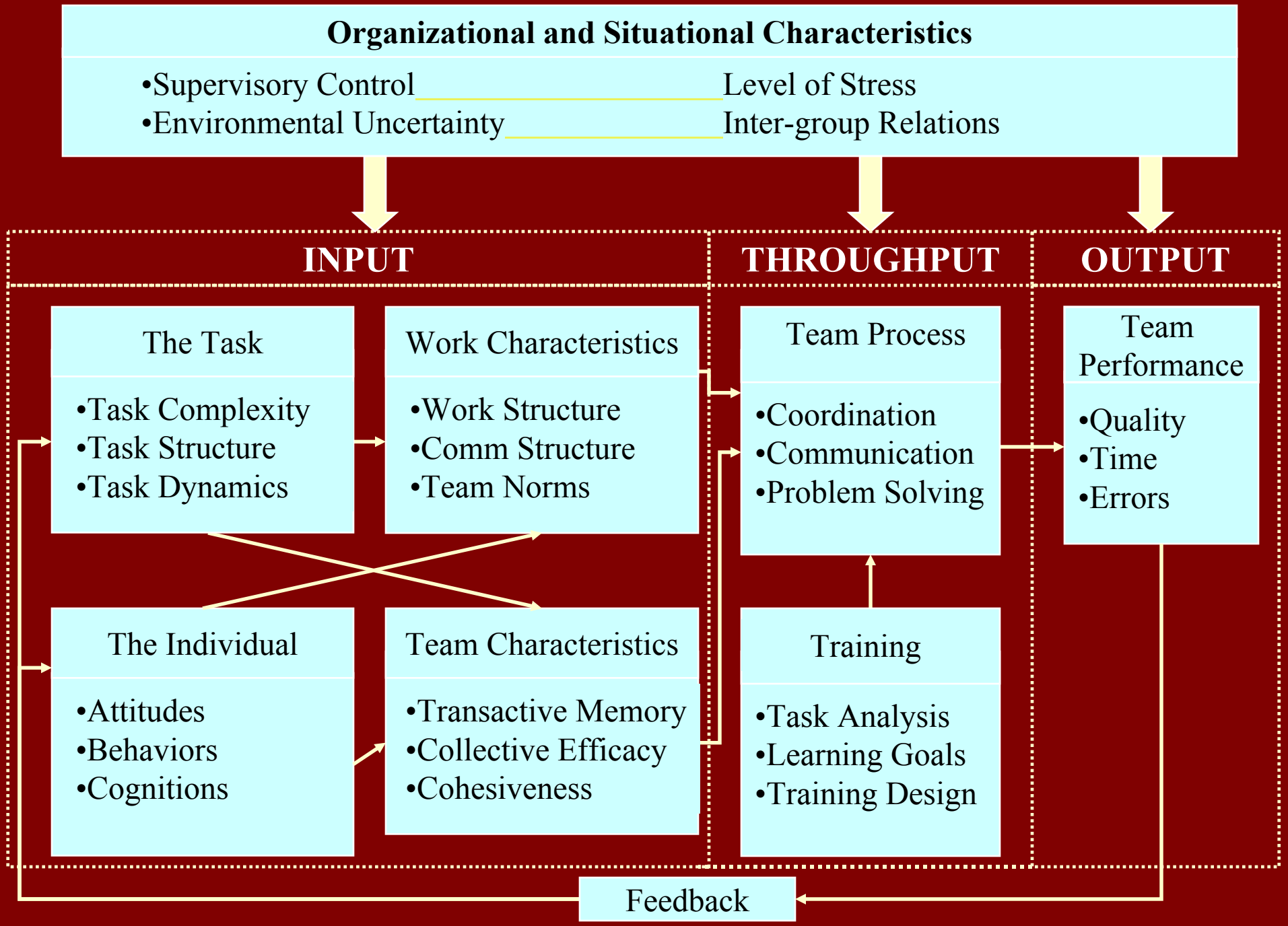
A Research Methodology and Initial Findings: Studying Combat Readiness and Stress Within Teams

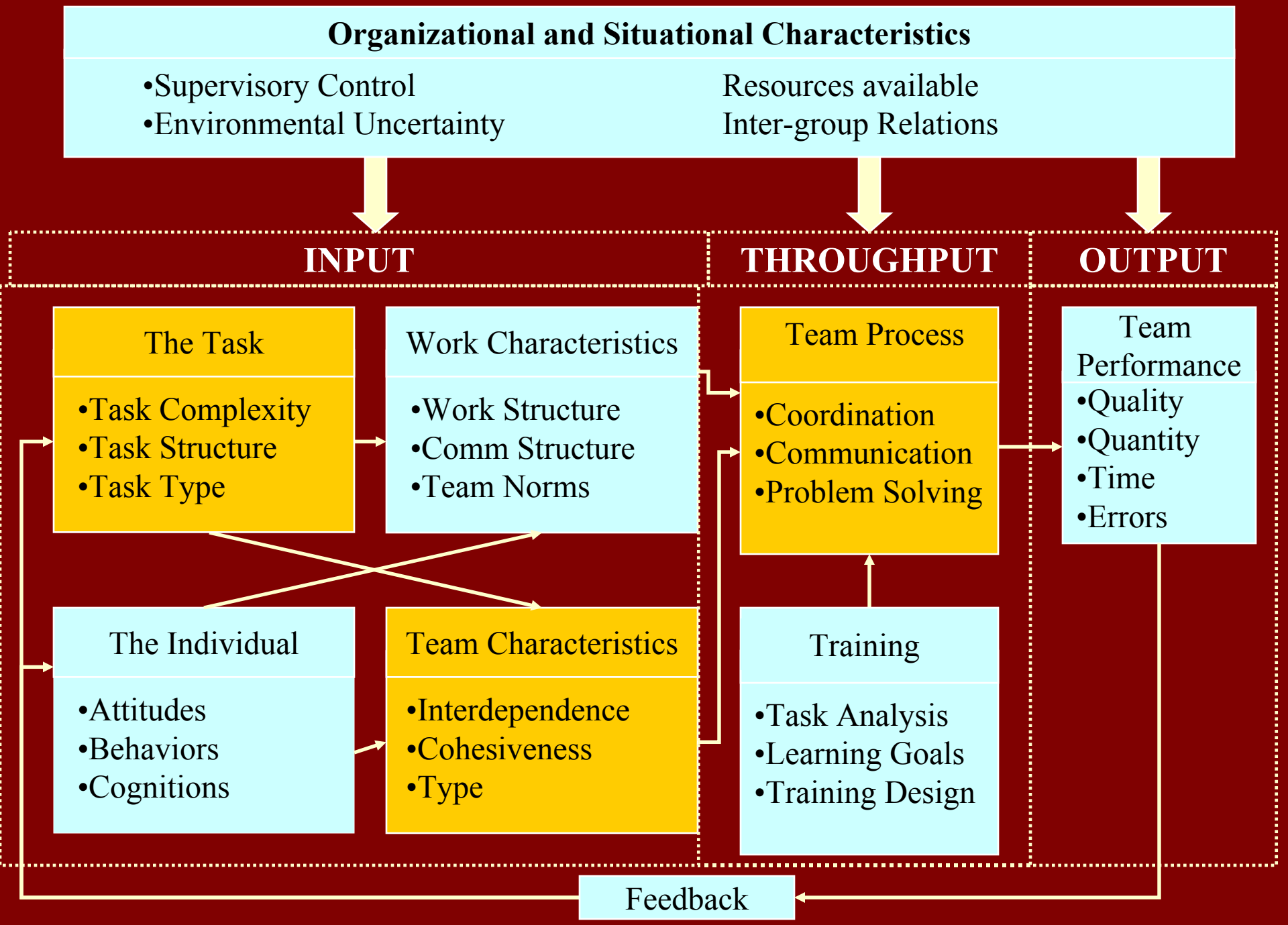
C. Shawn Burke, PhD, Clint Bowers, PhD., Eduardo Salas, PhD.,
Kia Asberg, Heather Priest, MS., and Laura Milham.

Institute for Simulation and Training
Team Performance Laboratory
University of Central Florida

A Word From Our Sponsor

- The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DOD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.


Organizational and Situational Characteristics
•Supervisory Control
•Environmental Uncertainty
Level of Stress
Inter-group Relations
INPUT
THROUGHPUT
OUTPUT
The Task
•Task Complexity
•Task Structure
•Task Dynamics
Work Characteristics
•Work Structure
•Comm Structure
•Team Norms
Team Process
•Coordination
•Communication
•Problem Solving
Team Performance
•Quality
•Time
•Errors
The Individual
•Attitudes
•Behaviors
•Cognitions
Team Characteristics
•Transactive Memory
•Collective Efficacy
•Cohesiveness
Training
•Task Analysis
•Learning Goals
•Training Design
Feedback


Organizational and Situational Characteristics
•Supervisory Control
•Environmental Uncertainty
Resources available
Inter-group Relations
INPUT
THROUGHPUT
OUTPUT
The Task
•Task Complexity
•Task Structure
•Task Type
Work Characteristics
•Work Structure
•Comm Structure
•Team Norms
Team Process
•Coordination
•Communication
•Problem Solving
Team Performance
•Quality
•Quantity
•Time
•Errors
The Individual
•Attitudes
•Behaviors
•Cognitions
Team Characteristics
•Interdependence
•Cohesiveness
•Type
Training
•Task Analysis
•Learning Goals
•Training Design
Feedback

Task Characteristics

Task Characteristics

- Complexity
- Structure
- Type

Stressors:

- Time Pressure
- Work Load
- Fatigue
- Ambiguity
- Noise

Team Characteristics

Types of teams in the literature:

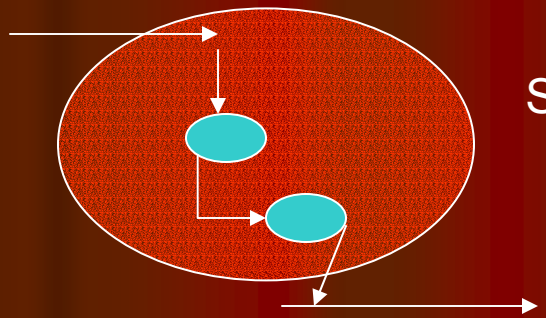
- **Distributed vs Co-located**
  - Virtual and/or real life distribution and location.
- **Interdependence**:
  Reciprocal, Sequential, Pooled, and Team (Saavedra, Earley & Van Dyne, 1993).
- Type of **cues** available:
  -Visual cues of team mates' actions, verbal cues from team mates, etc.
- **Performance feedback**
  - Team or individual.

Team Characteristics

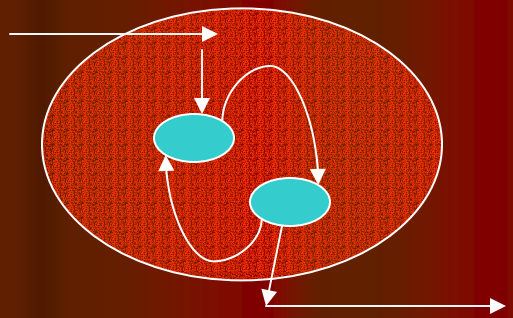
- Team Type
  - Distributed vs Co-located
- Interdependence
- Cohesiveness



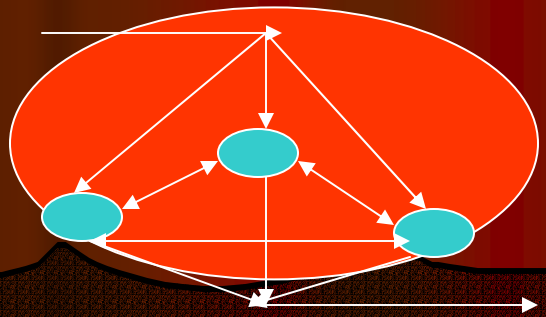
Task Interdependence



Sequential: One team member must act before another can act, but interaction is essential for effective performance.



Reciprocal: Team members have different roles and are often specialists with different expertise. They perform different parts of a task in a flexible order.



Team: Team members jointly diagnose, problem solve, and collaborate to complete a task. The task requires mutual interactions with team discretion to decide on particular course.

Team Processes

Team Processes

- Coordination
- Communication
- Problem Solving

- Explicit Coordination determine coordination strategies under low workload coordination.
- Implicit coordination takes over as workload increases.
  - Serfaty, Entin, Johnston, 1998
- A switch from explicit to implicit coordination is the primary adaptation mechanism that allowed teams to maintain and improve their performance under a high level of time pressure.
  - Entin & Serfaty, 1999

Adaptability

- Adaptability can be viewed as a sub component of communication and coordination and is an important part of effective team process and performance.

- Involves the adjustment of strategies by using information from the environment through compensatory behavior and reallocation of resources in stressful or novel situations. (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995).

Research Approach

- What the team performance literature includes and what it is lacking.

- Research Questions:

1. What type of stress effects generalize from individuals to teams and what effects are unique to teams?
2. What stress effects generalize across teams and what effects may be unique depending on team type?
3. How do teams adapt when circumstances change and how do they behave in response to stressors?

FLASHPOINT$_{TM}$ TESTBED

FLASHPOINT™ OBJECTIVES

- **Intelligence based** mission with **verbal communication** as the key to successful performance.
- **Two member distributed team**
- **Time pressure** is manipulated through **false feedback** (N+2).
- Objective is to **verify truthfulness** of obtained information regarding terrorist activity by locating objects, hostages and **identify enemy targets** within **3D setting**.
- Teams make a **decision** to release or detain information sources depending on their affiliation (terrorist or allied).

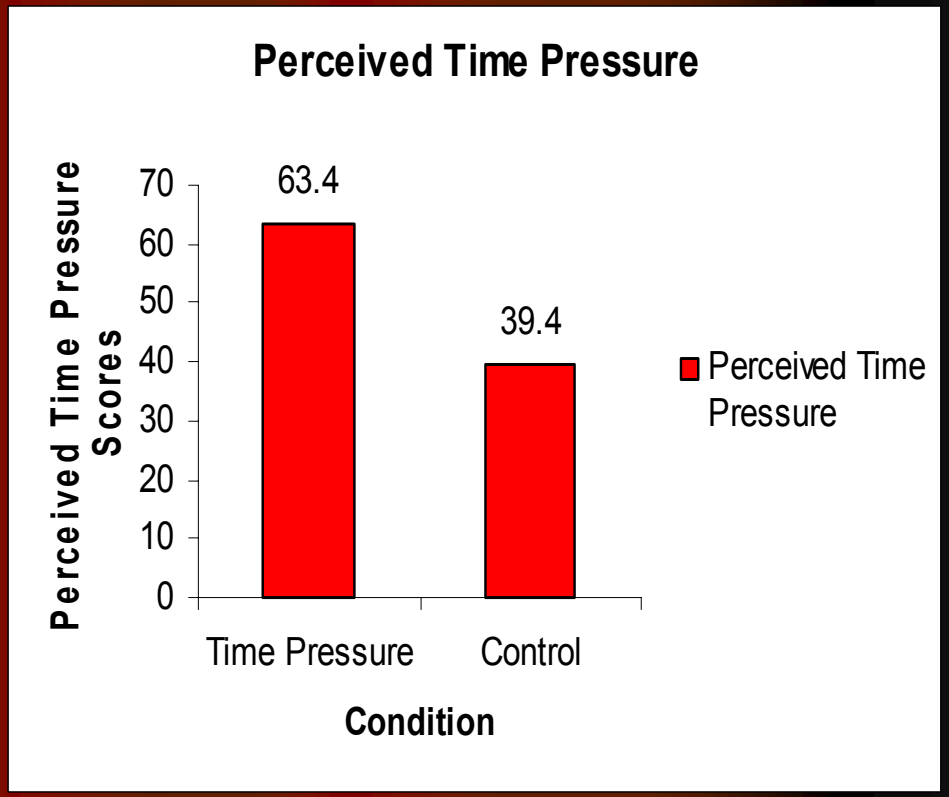


False Positive Error (FPE)	Verifying the truthfulness of a statement when it is false, e.g., circling an object or target that is not there.	How well team members communicate directions to the target and their position (through use of environmental cues, maps, etc).
False Negative Error (FNE)	Holding a statement as false when it is true, e.g., circling an object or target that is not there.	How well team members communicate directions to the target and their position (through use of environmental cues, maps, etc).
Time Error (TE)	Behavior violating the basic premise of "working efficiently" through the mission, e.g., "time wasting behavior" such as telling a team member to go to a location they have visited.	How much attention team members pay to the objective of finding the most time effective way to complete as many mini-objectives as possible.
Correct Identification – Object (CIO)	Correct verification or falsification of a specific individual statement in the cluster (e.g., circling "T" when statement is indeed true).	Team's ability to correctly convey information, navigate through the mission, and obtain objectives of the task.
Correct Identification – Cluster (CIC)	Correct verification or falsification of a cluster (e.g., writing a "T" next to cluster when at least 2 out of 3 individual statements are true)	Team's ability to correctly convey information, navigate through the mission, and obtain objectives of the task.

FLASHPOINT$_{TM}$ RESULTS

FLASHPOINT RESULTS:

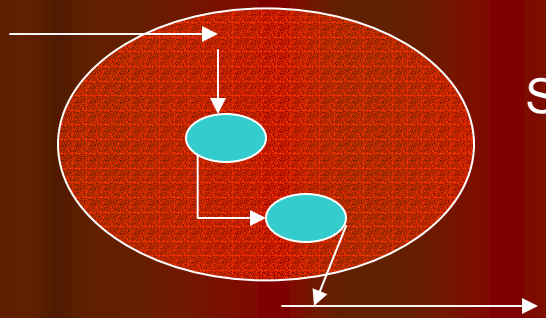
- 3 Pilot studies for scenario development (completed).
- Time Pressure study (N=20):
- Assessment of perceived time pressure using the NASA TLX has yielded significant group differences.
- Actual difference in performance between the two conditions was non-significant
- Communication analysis (ongoing).
- Study results by summer.



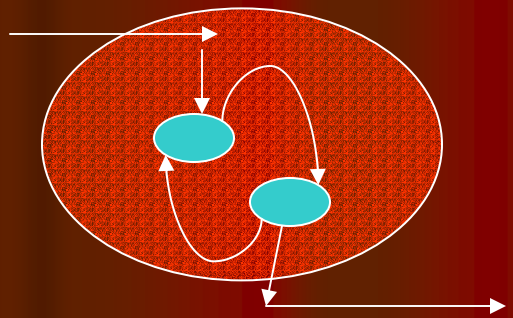
Team Types

Characteristics	Flashpoint™	DDD	Ghost Recon ™
Task Interdependence Saavedra, Earley, & Van Dyne, 1993	Sequential	Reciprocal	Team
Available cues of team mates' actions	Verbal Only	Verbal and Low Detail Visual	Detailed Visual and Verbal
Similar to:	Reconnaissance Team	Command and Control Team	Special Forces Combat Team

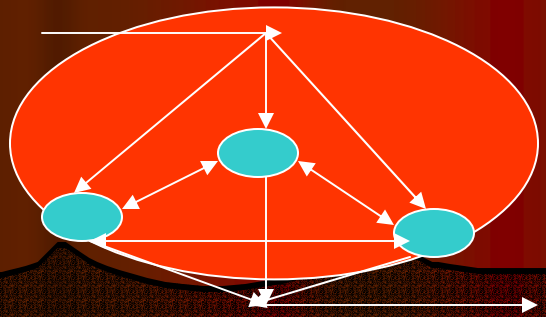
Task Interdependence



Sequential: One team member must act before another can act, but interaction is essential for effective performance.



Reciprocal: Team members have different roles and are often specialists with different expertise. They perform different parts of a task in a flexible order.



Team: Team members jointly diagnose, problem solve, and collaborate to complete a task. The task requires mutual interactions with team discretion to decide on particular course.

Workload Series

- Manipulation
- Measures
  - NASA-TLX
  - Dundee Stress State
  - Revised MSLQ
  - Communication Analysis
  - Team Process Analysis

FLASHPOINT$_{TM}$ TESTBED

FLASHPOINT™ OBJECTIVES

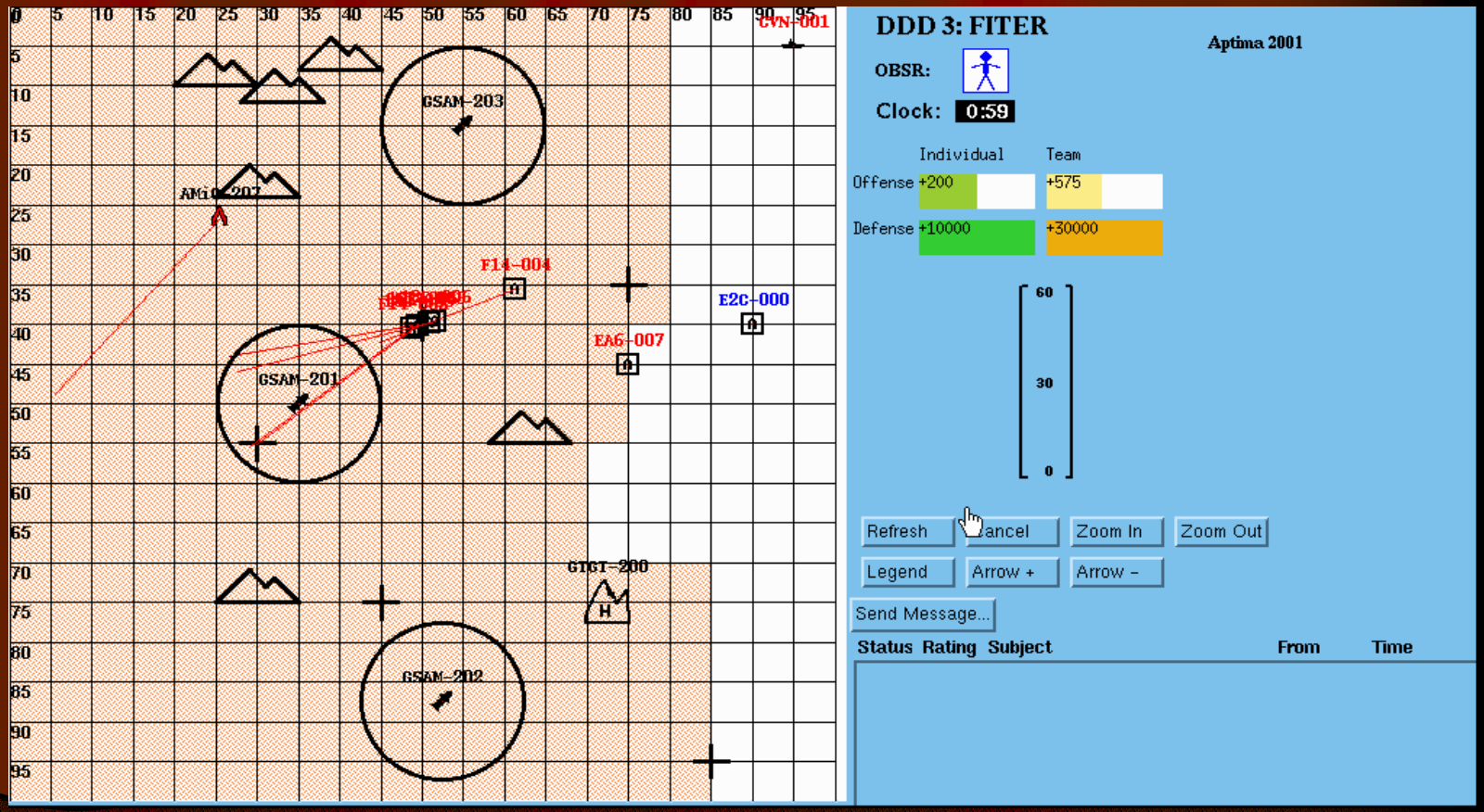
- **Intelligence based** mission with **verbal communication** as the key to successful performance.
- **Two member distributed team**
- **Workload** is manipulated through dual auditory task.
- Objective is to **verify truthfulness** of obtained information regarding terrorist activity by locating objects, hostages and **identify enemy targets** within **3D setting**.
- Teams make a **decision** to release or detain information sources depending on their affiliation (terrorist or allied).



DDD TESTBED

DDD OBJECTIVES:

Goal is to fly a **checkpoint mission** and drop bombs on a target at a specified time.



DDD TESTBED

DDD TASK:

Participant A: identified unknown aircraft and ground sites and monitored timing and dangerous cue patterns (late to a checkpoint; enemy aircraft approaching the team member); alerted team member of threatening events.

Participant B: flew checkpoint mission, avoided enemy air and ground threats, and destroyed high priority enemy aircraft.

DDD MISSION:

- Scenarios crafted with specific events to occur throughout the mission.

DEPENDENT MEASURES:

- Task performance was measured through the number of the team's aircraft that were destroyed, the number of enemies that were destroyed; and the time that each checkpoint was reached.

GHOST RECON™ TESTBED

GHOST RECON OBJECTIVES

- **Tactical based** mission with verbal communication and , coordinated team action serving as the key to successful performance.
- **Two member co-located team**
- **Workload** is manipulated via a dual auditory task requiring participants to respond with their location at specified points (30 second intervals).



GHOST RECON™ TESTBED

- **GHOST RECON™ MISSION:**
  - Goal is to complete **4 objectives** (e.g., gather intelligence, destroy a SAM site) within a 30 minute performance mission.
  - Subjects were randomly assigned to a **two member team** equipped with specific tools (i.e., weapons) necessary to complete the specific mission.
  - Outcome measures include:
    - Incidents of Fratricide
    - Communication
    - Objectives Completed
    - Mission Completion Time

Mission1:

1. Gather intelligence from SE house
2. Gather intelligence from NW house
3. Destroy SAM site
4. Get to extraction point



GHOST RECON™

- **GHOST RECON™**
  - Pilot studies have been conducted.
  - Pilot studies using the NASA-TLX to measure workload have yielded several trends in the predicted direction.
  - Pilot studies have indicated observable teamwork behaviors.
  - Significant outcomes…



Perception of Individual Workload

Scale	Low Workload	High Workload
Mental Demand	46.9	67.0
Frustration	37.9	66.0
Physical Demand	4.6	3.0
Temporal Demand	28.4	29.0
Performance	25.8	41.0
Effort	47.6	35.0

***N = 16 (individual level)**
8 teams of 2

Perceptions of Team Workload

Scale	Low Workload	High Workload
Mental Demand	46.7	47.2
Frustration	34.6	46.0
Physical Demand	9.5	4.0
Temporal Demand	17.0	33.0
Performance	22.3	47.0
Effort	56.3	61.7

*** N = 8 (aggregated to the team level)**

Where We Are Going…

- Analysis, Analysis, Analysis…..
- Research Questions:
  - Why would there be differential stress effects in teams with different characteristics?
    - What might the effects be? What about their differences influences the stress effects?
  - What are the differences in moderators and coping styles between teams with different characteristics?
    - Are there meta-behaviors that are universal across teams?
  - How do teams shift from operating in one situation to another (e.g., from attack to retreat) in a changing environment?
    - How do teams adapt?
- Objectives: Look at Teams, Adaptability, and Coordination more specifically.
  - Current literature approaches teams and training globally.

DEMONSTRATING A METHODOLOGY FOR ASSESSING COGNITIVE CAPACITIES ACROSS SENSORY SYSTEMS: M-SWAP

J. Christopher Brill, Mustapha Mouloua, Peter A. Hancock, Richard D. Gilson
University of Central Florida
Orlando, Florida
and
Robert S. Kennedy
RSK Assessments, Inc.
Orlando, Florida

This demonstration highlights the development of and applications of a methodology for assessing cognitive workload across three sensory systems: vision, audition, and touch. The methodology, called Multi-Sensory Workload Assessment Protocol (M-SWAP), is typically administered simultaneously with the primary task of interest. Participants are presented with a random series of visual, auditory, or vibrotactile signals and are asked to respond via a response box after receiving a certain number of these signals. Counting errors are scored in terms of both classic and fuzzy signal detection metrics (hits, misses, false alarms, and correct rejections), and the number of errors is indicative of the operator's cognitive-perceptual workload. The demand level of M-SWAP is scalable, which avoids traditional ceiling effects.

INTRODUCTION

Although a variety of workload measures exist, no extant measures are capable of assessing the capacities of cognitive resource pools across multiple sensory modalities and across various tasks (Meshkati, Hancock, Rahimi, & Dawes, 1995). The goal of the present effort was to develop a device and protocol for assessing the workload associated with complex tasks involving more than one sensory system. Five design criteria were established to guide the development of the workload assessment device and protocol. First, portability is essential to maximize the utility of the device across research scenarios, regardless of whether it is used in the field or in the laboratory. Second, the protocol should be independent of language requirements and cultural bias. Third, use of the device should not require special abilities such as mathematical computation or reading. Fourth, the protocol should be easy to learn, and there should be an absence of substantive or prolonged practice effects. Fifth, the device should be inexpensive to build and maintain to increase its accessibility. We suggest that the workload assessment device and protocol envisioned below meets all five of these criteria.

METHOD

Apparatus

The workload assessment device consists of a wearable display as an input to the least used sensory modality: touch. The system is comprised of three vibrotactile actuators (called tactors), a relay-based tactor controller board, a three-button response box, a portable 5-inch LCD screen, and a PC using proprietary software developed by the research team. The system is used for presenting a multi-sensory counting task for assessing attentional capacities while operators perform complex tasks (such as driving while using telematic devices including cell phones, in-car navigation systems, or even tuning a radio).

Procedures

The protocol for using the workload assessment device is generally the same regardless of which sensory modality (i.e., visual, auditory, or tactile) is targeted, but the following description depicts the visual component of the protocol. Participants are presented with a series of lights flashing randomly at three locations (left, center, or right) on a computer screen. They count the frequencies of lights (each four count), and respond by pressing one of three response buttons. The number of counting errors serves as a measure of workload, wherein low workload tasks should be associated with few counting errors and high workload tasks should be associated with greater frequencies of counting errors. The difficulty of the counting task is scalable depending upon the number of stimulus sites subjects are instructed to monitor (one site = low difficulty, two sites = moderate difficulty, etc.). The counting task for the other two sensory modalities is similar wherein subjects are counting from among either three loci of vibrotactile stimulation (left side of torso, midline, or right side of torso) or three auditory pitches (low, middle, and high). The counting task is an adaptation of the one used by Kennedy (1971), which itself is a derivation of one used by Jerison (1955, 1956).

DISCUSSION

Advantages of M-SWAP

Aside from meeting the five aforementioned design criteria, there are other significant advantages of the multi-sensory workload assessment protocol (hereafter referred to as "M-SWAP") over other established workload measures. For example, the NASA Task Load Index (NASA-TLX) is a commonly used subjective self-report workload measure in which operators rate items on several task dimensions such as mental demands, temporal demands, effort, and frustration (Hart & Staveland, 1988). It is limited in that it cannot be administered while someone is actually performing a task, as can M-SWAP. In addition, the NASA-*TLX* treats cognitive resources as a unitary pool (e.g., Broadbent, 1958), which restricts its utility for situations in which one wishes to determine the differential demands of a task on cognitive resources, especially if one's research is based upon theories purporting the existence of multiple resource pools of attention (cf., Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1984). In fact, assessing resources across and within modalities is a distinct advantage of M-SWAP. With M-SWAP, cognitive capacities can be assessed on-line (meaning while the operator performs the task) and in real-time, without the limitations of task-specificity associated with primary task measures (e.g., Linton's (1975) statistical workload assessment model (SWAM); North and Riley's (1988) Workload Index (W/INDEX). This real-time (or nearly real-time) assessment is essential for the design of adaptive technologies in which elements of a complex task might be automated through feedback if excessive workload levels are detected (Hancock & Chignell, 1988). Finally, the secondary task paradigm used by the M-SWAP can be used to determine the availability of resources across sensory modalities. These data can also be used for system design to maximize the likelihood of signal detection while performing complex tasks. For example, one might use the M-SWAP for confirmation of the availability of resources within a sensory modality that is theoretically untaxed by a task to aid in making design decisions (e.g., to what sensory modality should an alarm system target).

Continuing Research

Testing to establish the validity of the M-SWAP is continuing, but initial evaluation of the device and protocol in both a driving simulator and in an on-road vehicle suggests the M-SWAP is sensitive to different levels of workload. For example, while operating an automobile in light traffic participants committed an average of one counting error (visual) every four minutes (Mouloua, Rinalducci, Hancock, & Brill, 2003). When given the additional task of either operating a navigation system or talking on a cell phone, mean counting errors increased 2.5 and 4.5 per four-minute block, respectively. Pilot data indicates that participants' performance on M-SWAP is highly consistent across administrations ($\rho = .961, p < .03$).

ACKNOWLEDGEMENTS

This research was facilitated by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ARO under grant DAAD19-01-1-0621. P. A. Hancock, *Principle Investigator*.

The views expressed are those of the authors and may not reflect those of the sponsor.
Correspondence regarding this work should be directed to Chris Brill at jbrill@pegasus.cc.ucf.edu.

REFERENCES

Broadbent, D. (1958). *Perception and communication*. New York: Pergamon.

Hancock, P. A., & Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man, and Cybernetics, 18,* 647-658.

Hart, S., & Staveland, L. (1988). Development of the NASA-TLX: Results of empirical and theoretical research. In P. Hancock and N. Meshkati (Eds.), *Human mental workload.* Amsterdam: North-Holland.

Jerison, H. J. (1955). *Effect of a combination of noise and fatigue on a complex counting task* (Report Number WADC TR-55-360). Dayton, Ohio: Wright Air Development Center, Wright Patterson AFB.

Jerison, H. J. (1956). *Differential effects of noise and fatigue on a complex counting task* (Report Number WADC TR-55-359). Dayton, Ohio: Wright Air Development Center, Wright Patterson AFB.

Khaneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kennedy, R. S. (1971). A comparison of performance on visual and auditory monitoring tasks. *Human Factors, 13(2),* 93-97.

Linton, P. (1975). *VFA-STOL crew loading analysis* (NADC-57209-40). Warminster, PA: U.S. Naval Air Development Center.

Mouloua, M., Rinalducci, E., Hancock, P. A., & Brill, J. C. (2003). *Evaluating workload associated with telematic devices via a secondary task protocol.* Paper presented at the Second International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Park City, UT.

Meshkati, N., Hancock, P. A., Rahimi, M., & Dawes, S. M. (1995). Techniques in mental workload assessment. In J. R. Wilson and E. N. Corlett, (Eds.) *Evaluation of human work: A practical ergonomics methodology* (pp. 749-782). Taylor & Francis, Philadelphia, PA.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86,* 214-255.

North, R., & Riley, V. (1988). W/INDEX: A predictive model of operator workload. In G. MacMillian (Ed.), *Human performance models.* Orlando, FL: NATO AGARD Symposium.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63-101). Hillsdale, NJ: Erlbaum.

THE EFFECT OF AFFECT:
THE HEDONOMIC EVALUATION OF HUMAN-COMPUTER INTERACTION

Lauren Murphy, Kay Stanney, and Peter A. Hancock

Department of Psychology and Institute of Simulation and Training, and
Department of Industrial Engineering and Management Systems
University of Central Florida
Orlando, FL 32816

We evaluated the use of color in a human-computer interface to investigate how affective cues influence usability judgments. Yellow and green, were found to enhance performance on a cognitive flexibility task as compared to white and gray. Such an outcome suggests that color promotes positive affect. The aesthetically designed interface that promoted positive affect was also found to enhance perceptions of interface usability. This relationship, however, was only found when underlying usability was effective. When usability was poor, aesthetics made users more cognizant of the usability weaknesses. The results suggest that designers consider incorporating color into designs to help elicit positive emotion in users but that they must be aware that incorporation of aesthetics comes with the potential cost of disposing users to be more discerning of usability.

INTRODUCTION

Affects have significant effects which need elucidation so that a full portraiture of human-machine interaction can be generated. All operational domains have experienced a significant change as workload has shifted from largely physical to largely cognitive composition. The transition in work has changed the 'form' of the demand but not the 'perception' of the demand, which it is still viewed as 'work' and thus not enjoyable (Hancock, 1997). In the evolution of human-technology interaction a movement is now emerging that seeks to promote pleasure through design (Jordan, 2000; Helander, 2002). Until recently the design of artifacts was independent of the explicit consideration for aesthetics and with a few odd exceptions, ubiquitously placed usability goals ahead of enjoyable interaction (Norman, 1990). Today's advances in technology enable the flexibility to achieve usability goals as well as promote pleasure in users (Jordan & Servaes, 1995). Hedonomics, a term coined by Hancock (2002), proposes to dissolve what we conceive as the adverse or unpleasant attributes of work by aiming to facilitate the design of work, and all aspects of it, to be intrinsically enjoyable.

In the realm of Hedonomic research, the present study investigated the relationship between aesthetics and usability by incorporating color into design with the intent to promote positive affect in the user and by concomitant measures the user's perceived usability of the system.

Attribution Theory

The relationship between aesthetics and perceived usability is congruent with the social phenomenon of inferring personality traits from physical attractiveness. Dion, Berscheid, and Walster (1972) found that people that were viewed as physically attractive were assumed to possess more socially attractive traits than those that were viewed as unattractive. A possible explanation for this phenomenon is the halo effect, which proposes that the most obvious or salient characteristic (in this case, attractiveness) is perceived first and tends to bias perceptions and inferences that come after. Furthermore, social psychology research reveals that initial perceptions persevere even after presentation of contrary evidence (Gilbert, Krull, & Malone, 1990). Based on this halo effect, users may attribute more desirable traits (such as ease of use, ease of learning) to interfaces that are designed to be aesthetically pleasing compared to interfaces that are not aesthetically pleasing. Users may even continue to attribute desirable traits to aesthetically pleasing interfaces even after they are presented with evidence to be contrary.

Aesthetics is a concept that is defined subjectively and therefore difficult to manipulate systematically in an experimental setting. There

are clear usability guidelines in interface design but no clear aesthetic guidelines. A rule of thumb is that aesthetics should promote pleasure by eliciting positive affect in the human (Hancock, 2002; Norman 2002). According to Norman one way of viewing aesthetics can be through color. Switching from black and white displays to color displays doesn't have an obvious affect on the usability of the display but does have an obvious effect on the aesthetics of the display. Most people prefer color displays. Color may have some emotional affect on the user rendering them to prefer a color to black and white display.

Positive affect and cognitive flexibility

While color has still to be shown to directly relate to affective state (Sinclair, Moore, Lavis, & Soldat, 2002), it has already been shown to differentially affect information-processing strategy. More specifically, negative and neutral affective colors have been shown to lead to more systematic, discerning processing, while positive affective colors render a more accepting, indiscriminate processing approach. Thus, negative affect may lead to greater cognitive flexibility, where an individual perceives and interprets information from multiple perspectives and in greater detail. Yet, cognitive flexibility has been shown to increase with the presentation of a stimulus used to elicit positive affect in a person (Isen, 2000; Isen et al., 1987). Several studies (Estrada et al, 1994; Isen et al, 1987) have found that promoting positive affect improves cognitive flexibility.

This presents a conundrum. Will positive color in a design lead people to process information in less detail and thus potentially gloss over usability weaknesses or will it lead to greater cognitive flexibility and a more discerning assessment of usability? If the former is true, products should liberally incorporate color as it could mask usability weaknesses. If the latter is true, incorporating color could render usability of utmost importance, as users would readily perceive any weaknesses. The objective of the current study was to resolve this question by determining how affective color cues influence usability judgments.

METHOD

Participants

Participants were 12 undergraduate students, between the ages of 19 to 26, seven males and five females, enrolled in a psychology course, and who participated in the experiment for extra credit. Participants were screened to ensure color vision capacity.

Task

Participants were asked to compute 15 multiplicative calculations on one of two computer-programmed calculator designs. One of the designs had a white background and gray keys, the other had a yellow background and green keys (see Terwogt & Hoeksma, 1995; Peretti, 1974). Thereafter the participants were asked to fill out an evaluation form of the calculator interface using a modified satisfaction questionnaire (Shneiderman, 1991). The final task was to perform a Remote Associate Test (RAT), which can be used to assess cognitive flexibility (Mednick. 1962).

Procedure

Participants were randomly assigned into either the experimental group or the control group. Participants in the experimental group were presented with the yellow and green calculator display (on the computer interface) twice, one display with high usability (no delay in feedback) and one with low usability (75 ms delay in feedback). Each time the calculator display was presented, participants were asked to perform a multiplication task. These tasks consisted of multiplying a string of 15 different numbers. In the second multiplication task, the numbers changed but still remained as a string of 15 different numbers. After the participants completed their multiplication task they were asked to rate the calculator display's general usability via the satisfaction questionnaire. Then participants were asked to complete the RAT.

Experimental Design and Manipulation

The experiment used a 2x2 mixed design. Aesthetics level (high [color] vs. low [no color]) of the calculator interface was the between-subject variable. The within-subject variable was interface usability level (high [no delay in feedback] vs. low [75 ms delay in feedback]). Tractinsky, Katz, and Ikar (1999) used similar manipulations. Participants were shown either the color or the no-color interface, once with high usability (no delay) and once with low usability (75 ms delay), with the order of

feedback presentation being randomized among participants. After each exposure to the calculator, participants rated the usability of the calculator display and completed five items from the RAT. Thus, the dependent variables of the study include "ease of use" or the user's perceived usability of the system, and "positive affect" or degree of cognitive flexibility.

RESULTS

The first objective was to investigate if color elicits greater cognitive flexibility in users. If color is an affective stimulus in interface design then participants from the experimental group (those receiving the color design) should score higher on the RAT than participants in the control group (those receiving the no color design). ANOVA results revealed a significant effect of color on the participants' RAT scores, $F(1,9)=4.285$, $p=.05$. Participants that received the colored calculator displays ($M=51$, $SD=3.27$) performed significantly better on the RAT compared to the participants that received the non-color displays ($M=47$, $SD=2.79$). These data indicate that the incorporation of color into a design can enhance cognitive flexibility.

The next objective was to determine how greater cognitive flexibility influenced usability judgments. ANOVA results revealed no significant main effect for aesthetics (i.e., color), but a significant main effect for usability level (see Figure 1) on the participants' perceived usability scores ($F(1,9)=9.085$, $p=.01$), and a significant interaction between aesthetics and usability ($F(1,9)=7.279$, $p=.01$), see Figure 2.

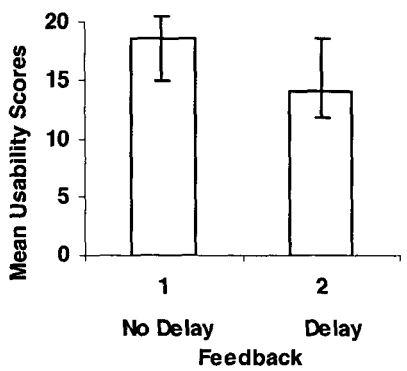


Figure 1. *A graphic display of the perceived usability means for the IV-delay in feedback, where '1' indicates no delay in feedback and '2' indicates delay in feedback.*

Pairwise comparisons indicated that participants rated the color calculator with a delay in feedback (CD) significantly ($P=.05$) lower in usability than all other displays. The color calculators with no delay (CND) in feedback were rated significantly ($p=.05$) higher in usability than all other displays. There was no significant difference in perceived usability between the delay in feedback (NCD) and no delay in feedback (NCND) conditions for the non-color calculator (see Figure 2).

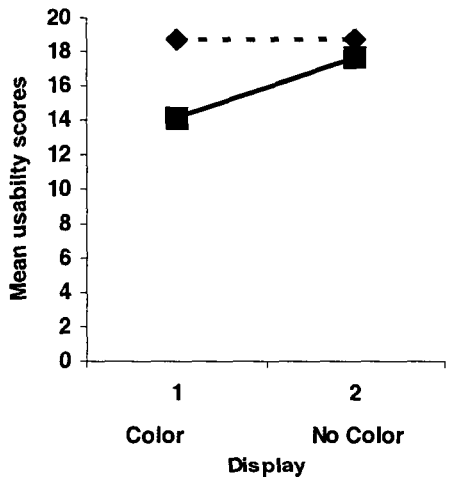


Figure 2. *A graphic representation of the means for the interaction between aesthetics (1 = color and 2 = no color display) and usability level (dashed line = no delay in feedback and solid line = delay in feedback.*

DISCUSSION AND CONCLUSIONS

This study has contributed two new findings toward the promotion of pleasure in human-technology interaction. First, aesthetically pleasing designs may only promote pleasure in the user when coupled with good usability criteria. As revealed in Figure 2, despite the aesthetic manipulation, color displays were not pervasively rated higher in usability than non-color displays when the two designs (both with and without delay) were compared. Interestingly, it appears that participants became the most frustrated with poor usability when a color display was used as compared to a non-colored display. This finding contradicts the hypotheses and previous findings that reported an increase in perceived usability of aesthetically pleasing designs regardless of their actual usability (Kurosu & Kashimura, 1995; Tractinsky, Katz,

& Ikar, 1999). Generally, participants preferred the color display with no delay over the other displays until the delay of feedback was incorporated into the design, then participants preferred the non-colored display. The results suggest that aesthetically designed interfaces may enhance perceptions of usability only when underlying usability is strong. When usability is weak, aesthetics may make users more perceptive of usability weaknesses. Thus, if a product has known usability flaws, it may be best to minimize product aesthetics. Second, colors such as yellow and green were shown to enhance performance on cognitive flexibility tasks (i.e., RAT), suggesting that they have the potential to promote positive affect in users.

Therefore, designers should consider incorporating color into designs to help elicit positive emotion in the user. Designers must be aware, however, that incorporation of aesthetics comes with the potential cost of disposing users to be more discerning of usability. It is thus essential that designers fully understand the affective consequences of their designs. With today's advances in technology we now have the flexibility to achieve both usability goals as well as promote pleasure in the user. The findings from this study confirm the importance of both in good human-technology design.

ACKNOWLEDGEMENT

This work was facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, Dr. P.A. Hancock, *Principal Investigator*. The assistance of the Department of Psychology along with Pär Axelsson, Jon Berglin, Mattias Klefbeck, and Aaron Pepe from the Department of Industrial Engineering at University of Central Florida are also gratefully acknowledged. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

REFERENCES

Adams, F., & Osgood, C. E. (1973). A cross cultural study of the affective meaning of color. *Journal of Cross-Cultural Psychology,* **4**, 135-156.

Birren, F. (1978). *Color and human response*: New York: Van Nostrand Reinhold.

Dion, K., Berscheid, E., & Walster, E., (1972). What is beautiful is good. *Journal of Personality and Social Psychology,* **24**(3), 285-290.

Eyesnck, H.J. (1941). A critical and experimental study of color preference. *American Journal of Psychology,* **45**, 385-394.

Frank, G. (1976). On the validity of a hypothesis derived from the Rorschach: The relationship between color and affect. *Perceptual & Motor Skills,* **43**, 411-427.

Gilbert, D.T., Krull, D.S., & Malone, P.S. (1990). Unbelieving the unbelievable: some problems in the rejection of false information. *Journal of Personality and Social Psychology,* **59** (4), 601-613.

Hancock, P.A. (1997). *Essays on the future of human-machine systems.* MN: Banta Information Services.

Hancock, P.A. (2002). Hedonomics: The science and design of enjoyable human-technology interaction. *Report of the MURI Research Laboratory, MURI-OPUS,* IR-01-02, Orlando, FL.

Helander, M.G. (2002). Hedonomics-affective human factors design. *Proceedings of the Human Factors and Ergonomics Society,* **46**, *725-729.*

Isen, A.M. (2000). Positive affect and decision making. In M. Lewis and J.M. Haviland-Jones(Eds.), *Handbook of emotions* (2nd ed., pp. 417-435).

Isen, A.M., Daubman, K.A., & Nowicki, G.P. (1987). Positive affect facilitates creative problem solving. *Medical Decision Making,* **11***(3)*, 1122-1131.

Jordan, P.W., & Servaes, M. (1995). *Pleasure in product use: beyond usability.* In S. Robertson (Ed), *Contemporary Ergonomics.* (London: Taylor Francis). (pp. 341-346).

Jordan, P.W. (2000). The four pleasures- A framework for pleasures in design. In Jordan, P.W. (Ed), *Proceedings of Conference on Pleasure Based Human Design*, Groningen, The Netherlands: Philips Design.

Lewis, M., & Haviland, J.M. (1993). *Handbook of emotions.* NY: Guilford Press.

Mednick, S.A. (1962). The associative basis of the creative process. *Psychological Review,* **69**, 220-232.

Norman, D.A. (1990). *The design of everyday things.* New York: Doubleday/Currency.

Norman, D.A. (2002). Emotion and design: Attractive things work better. *Interactions Magazine,* **9**, (4), 36-42.

Peretti, P.O. (1974). Color-mood associations in young adults. *Perceptual and Motor Skills,* **39**, 715-718.

Terwogt, M.M., & Hoeksma, J.B. (1995). Colors and emotions: preferences and combinations. *Journal of General Psychology,* **122***(1),* 5-17.

Tractinsky, N., Katz, A.S., Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers,* **13**, 127-145.

M-SWAP: On-line Workload Assessment in Aviation

P. A. Hancock [1], J. C. Brill [1], M. Mouloua [1], R. D. Gilson [1], & R. S. Kennedy [2]
[1] University of Central Florida, [2] RSK Assessments, Inc.

Aviation environments are often fraught with excessive workload demands (Kantowitz & Campbell, 1996) and as a consequence a number of studies have examined issues of cognitive workload in aviation using techniques such as physiological measures (e.g., Wilson, 2002; Veltman, 2002; Bonner 2002; Backs, 1995) and subjective self-report measures including techniques such as the NASA-TLX (e.g., Moroney, 1999; Brookings, Wilson, & Swain, 1996). These techniques have several limitations. For example, physiological measures offer global assessments of workload, but they do little to parse the differential demands of tasks on sensory systems. Measures based upon retrospective self-report, such as the NASA-TLX (Hart & Staveland, 1988), are problematic as online workload assessment methods by their very nature. Lack of online workload assessment effectively limits their utility for use with adaptive automation systems. The goal of the present effort was to develop a methodology for online workload assessment that differentiates between the cognitive-attentional demands imposed by tasks across or within multiple sensory modalities (i.e., visual, auditory, and tactile). The Multi-Sensory Workload Assessment Protocol (M-SWAP) is an adaptation and elaboration of a counting task originally used by Kennedy (1971). M-SWAP involves randomly presenting participants with a series of visual, auditory, or tactile signals while an operator simultaneously performs the primary task of interest. Participants count the frequencies with which targets are presented and they respond by pressing a button. The number of counting errors serves as a measure of workload, wherein low workload tasks should be associated with few counting errors and high workload tasks should be associated with greater frequencies of counting errors. Data is analyzed in terms of traditional Signal Detection (SDT) metrics (hits, misses, etc.) and more recent Fuzzy Signal Detection (FSDT) measures (fuzzy false alarms, etc). Thus M-SWAP has the potential of providing very near real-time workload assessment that could be integrated into a cockpit adaptive automation system to regulate the cognitive demands on pilots when excessive workload is detected. A demonstration of this system will accompany the text presentation.

SKILLS, RULES, AND KNOWLEDGE: AN EXPERIMENTAL TEST OF PERFORMANCE DECREMENTS AS A FUNCTION OF STRESSOR EXPOSURE *

J. L. Weaver, K. A. Bradley, P. A. Hancock, J. L. Szalma, and J. A. Helmick
University of Central Florida

An important aspect of soldier effectiveness is the ability to perform multiple tasks successfully, often times under stressful conditions. However, it has been found that performance on a secondary task degrades when multiple tasks are being performed simultaneously and that the presence of stress has been shown to be associated with decrements in performance (Matthews, Davies, Westerman, & Stammers, 2000). Consequently, it is important to understand how the presence of stress impacts complex task performance. The purpose of the current study is to investigate the effects of various stressors on multiple task performance. Specifically, the tasks of the Multi-attribute Task Battery (MATB; Comstock & Arnegard, 1992) classified as skill, rule, or knowledge based (Rasmussen, 1983), will be investigated in terms of the order in which they will be shed, or degrade in terms of performance, as a function of exposure to various stressors.

Rasmussen (1983) classified human performance into three categories. *Skill-based* behaviors are routine activities conducted automatically and do not require conscious allocation of attentional resources. Behaviors are skill-based when human performance is determined by stored, preprogrammed patterns of instructions. Behaviors classified as *rule-based* are activities controlled by a set of stored rules or procedures (Reason, 1990). The distinction between behaviors classified as skill-based and rule-based depends on the attention and training level of each individual. Skill-based behaviors typically progress without conscious attention and the individual is seldom able to verbalize how performance on the behavior is controlled or explain the information on which their performance is based (Rasmussen, 1983). Performance of rule-based behaviors is typically based on specific ability and the rules for performance can be verbalized. *Knowledge-based* behaviors are those in which stored rules no longer apply and a novel situation is presented for which a plan must be developed to solve a problem. In contrast to set rules, plans are often required to be changed based on the situation. Attentional resources must be allocated to the behavior and, therefore, the performance of knowledge-based behaviors is goal-controlled (Rasmussen, 1983).

As mentioned, an integral component of Rasmussen's human performance model is attention. Two current theories of stress incorporate the importance of the delegation of limited attentional resources in response to stressors, each in a slightly different manner. Lazarus and Folkman's (1984) transactional model of stress places emphasis on the interaction between the person and environmental demands. Lazarus and Folkman define stress as the relationship between the environment and the person that the person perceives as exceeding or taxing his or her resources and endangering his or her well being. This model of stress is composed of four primary components that are theorized to proceed in a sequence. Behavioral, performance, subjective, and physiological consequences are then postulated to result as outcomes of this sequential process.

*This research was facilitated by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Hancock and Warm (1989) propose a dynamic model of stress in which a normative zone is surrounded by a comfort zone. It is within this area that a system or individual is believed to function at a normal target value. This zone resides between hypo and hyperstress regions indicating that a certain level of stress is necessary to operate in a normative fashion. When stress levels move closer to either hypo or hyperstress, the psychological maximal adaptability zone is pressed to its limit and eventually declines rapidly to minimal adaptability. The failure point for physiological maximal adaptability occurs further into hypo or hyperstress but represents a more serious danger. Performance decrements are postulated to occur as stress levels move toward hypo or hyperstress, likely due to persons' inability to return to a normal state because of the costliness of the resources employed to remedy the problem.

In order to investigate the effects of stress on multiple task performance, the Multi-attribute Task Battery (MATB; Comstock & Arnegard, 1992) will be used. The MATB is a computer-based task designed to evaluate operator performance and workload. The MATB requires the simultaneous performance of monitoring, dynamic resource management, and tracking tasks. The simultaneous performance of multiple tasks is a central feature of the MATB and it is this feature that is consistent with most operational systems and thus makes the task useful for our purposes as a research platform. Each of the MATB tasks can be considered to correspond with a specific performance category in relation to Rasmussen's human performance typology. The MATB tracking task requires the manual operation of a joystick in order to keep a target in the center of a tracking window. This tracking task is performed based on a relatively automated pattern of behaviors that do not require conscious attention and the task is therefore classified as a skill-based behavior. The monitoring task requires the resetting of a gauge once it has been detected as operating out of a specified range. Specific actions are taken in response to each gauge in accord with an *a priori* determined set of rules. As such, this task is classified as a rule-based behavior. The resource management task requires the activation of "pumps" in order to fill two tanks and also requires the maintenance of certain levels within them. This dynamic task has no predetermined rules for specific operation and therefore requires conscious processing due to the ongoing problem solving nature of the task.

The current study uses a modified version of the original MATB to investigate the effects of various stressors (i.e., noise, time pressure, ambiguity, workload, and threat) on multi-task performance. Specifically of interest is the order and degree to which performance degrades on each task based on their classification as skill-based, rule-based, or knowledge-based. The task labels on the screen of the MATB have been modified to resemble duties associated with military tank operation. In the first phase of the experiment, an aversive noise delivered through headphones at 85dB in random intervals will be presented as the stressor. Performance data will be collected along with data from the NASA TLX, an instrument that measures perceived workload. Data will also be collected from the Dundee Stress State Questionnaire (Matthews, Joyner, Gilliland, Campbell, Huggins, & Falconer, 1999) and the State Trait Anxiety Inventory (Spielberger, Gorusch, & Lushene, 1970).

The development of hypotheses regarding the current study is done so in congruence with the aforementioned notion that one must draw from a limited pool of resources in order to respond to conditions of stress. Specifically, we hypothesize that those tasks requiring more attentional resources will be shed, or degrade in terms of performance, first in response to a stressor,

whereas tasks more routinized will remain relatively undisturbed due to the lack of necessity for attentional resources. Specifically, it is hypothesized that the resource management task classified as knowledge-based will be shed first in response to noise, followed by the monitoring task classified as rule-based, and the skill-based tracking task will be the last to be shed.

References

Comstock, J. L., & Arnegard, R. J. (1992). *The Multi-attribute Task Battery for human operator workload and strategic behavior research*. Technical Report 104174. Hampton, VA: NASA Langley Research Center.

Hancock, P. A., & Warm. J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519-537.

Lazarus, R. S., & Folkman, S. (1984). *Stress appraisal and coping*. New York: Springer.

Matthews, G., Davies, D. R., Westerman, S. J., & Stammers, R. B. (2000). *Human performance: cognition, stress, and individual differences*. Philadelphia: Taylor and Francis.

Matthews, G. Joyner, L., Gilliland, K., Campbell, S. Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (vol 7, pp. 335-350).

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics, SMC 13,* 257-266.

Reason, J. (1990). *Human error.* Cambridge: Cambridge University Press.

Spielberger, C. D., Gorusch, R. L., & Lushene, R. E. (1970). *Manual for the state-trait anxiety questionnaire.* Palo Alto: Consulting Psychologists Press.

EFFECTS OF TASK DURATION AND TYPE OF AUTOMATION SUPPORT ON HUMAN PERFORMANCE AND STRESS IN A SIMULATED BATTLEFIELD ENGAGEMENT TASK

Kathleen McGarry, Ericka Rovira, & Raja Parasuraman
Cognitive Science Laboratory
The Catholic University of America

Automation that is meant to assist the operator by reducing stress and workload may actually increase both. Automation can be imperfect, and as a result may add to uncertainty and degrade performance when it fails. In a complex and dynamic task environment, this may lead to increased stress, which in turn may be associated with inefficient use of automation. The present study examined the relative benefits and costs of imperfect information and decision automation in a simulated battlefield engagement task. Performance and stress were evaluated for short- (10 minutes) and long-durations (30 minutes) of the task. The cost of unreliable aiding--a decrease in accuracy of appropriate target engagement selection--was greater for the decision automation than for information automation, irrespective of task length. Unreliability also slowed response times more for decision than for information automation, but only when the task was long. Subjective reports of stress also increased in the longer automated conditions. The results show that automated decision aids can improve performance when reliable. However, the price of imperfect decision automation is reduced accuracy and increased stress compared to information automation, particularly when the task must be performed for long periods.

INTRODUCTION

Current military operations involve a degree of complexity that mandates the use of automated decision-aiding tools to assist the human operator. Automation support is necessary because soldiers and commanders (e.g, in command and control) are required to make complex decisions under time stress, in the midst of information overload, and yet often with incomplete information. At the same time, there is increasing pressure in battlefield engagements to reduce the cycle time between target acquisition and response ("sensor to shooter" time), further necessitating automation. However, automation reliability cannot be assured in most domains, particularly in dynamic, uncertain environments such as command and control.

Automation can lead to both benefits and costs in human and system performance. A large body of research has shown that whereas automation generally enhances performance, reduces operator workload, and improves system efficiency, subtle costs can also be incurred, leading to a loss of coordination between human and automated agents. Over-reliance, mistrust, loss of operator skill, and unbalanced workload are among the issues in human-automation interaction that have been studied (Parasuraman & Riley, 1997). Stress is another issue facing those in combat situations. Stress can impact decision making strategies utilized by those in the situation, creating the possibility of incorrect assessment of information, leading to decisions that may not be the best course of action to pursue at that time (Hancock & Mortimer, 2002). Further, performing multiple tasks can itself be a source of stress (Matthews, Campbell and Falconer, 2001).

Automation can support the battlefield commander in many ways. It can (1) collect and present the multiple information sources in some organized manner. Automation can also (2) assist the human operator in high-stress, high-risk environments by integrating the multiple information sources and analyzing them for possible courses of action, allowing the operator to bypass a step in the decision-making process, and thereby possibly reducing stress. In addition, automation can (3) complete the information analysis and present the best course of action to the operator, or in effect make the decision for the human. Finally (4), the action could itself be automated. These four types of automation support have been termed information acquisition, information analysis, decision, and action automation, respectively (Parasuraman, Sheridan, & Wickens, 2000). The first two and the last two types can be grouped together and referred to as *information* and *decision* automation.

Following an early study by Crocoll and Coury (1990), a number of recent studies have compared the effects of these two forms of automation (Galster, Bolia, Roe, & Parasuraman, 2001; Galster, Bolia, & Parasuraman, 2002; Rovira, McGarry, & Parasuraman, 2002a; Rovira, Zinni, & Parasuraman, 2002b; Sarter & Schroeder, 2001; Wickens & Xu, 2002). A major finding is that, when reliable, both information and decision automation can enhance operator performance, but that unreliable automation can have a greater detrimental effect for decision than for information automation. Rovira et al. (2002a) also specifically found that for a simulated battlefield engagement task, reliable decision automation significantly reduced response time compared to information automation. Thus, decision automation significantly reduced the sensor-to-shooter time. However, this benefit was counteracted by

slower and less accurate engagement decisions when the automation was unreliable. There are two possible explanations for this pattern of results. The first is that operators become over-reliant on the automation when it provides decision and action choices for them, and do not check the underlying information sources as closely as when only information automation is provided (Crocoll & Coury, 1990). Alternatively, users may have insufficient experience of using the different information sources with decision automation, and may require greater time exposure to such automation. Perhaps the operator's mental model of the automation takes longer to build up when decisions are made for the operator. If so, the cost of unreliable decision automation might be mitigated or eliminated with longer task duration, and exposure to decision automation. We evaluated this possibility in the present study by comparing the effects of short (10 minute) and long (30 minute) task durations with either information or decision automation support in the same battlefield engagement task developed by Rovira et al. (2002a). Moreover, because longer task durations can be fatiguing and can reduce arousal, we also examined the effects of type of automation support and automation reliability on subjective measures of stress.

METHODS

Participants

Twenty volunteers aged 18-31 years (M=21.85, SD=3.81) were paid $15/hour for their participation.

Apparatus

A PC-based low-fidelity software simulation of a sensor-to-shooter targeting system (STS) was modified and used. A mouse was used as the input device.

Task and Procedure

The participants were required to identify the most dangerous enemy unit and to select a friendly unit to engage in combat with the enemy. Participants were given 6 seconds to make their selection. A 2 (automation type) x 2 (task duration) within-subjects design was used. Two types of automation were used: priority listing and recommendation automation. The priority listing automation provided participants with a complete list of all possible enemy-friendly combat engagements. The list gave distances between the friendly and enemy units, in a prioritized order, with the best choice first and the worst choice last. This corresponds to information automation because the distances to the headquarter units were not given to the participants, but were required to complete the decision process. The recommendation automation suggested the best option, but no distance information was given. This automation corresponds to decision automation, as it gave a suggested response. Figure 1 shows the Sensor-to-Shooter interface without automation support.

The automation support was 80% reliable across all automated trials. The participants were told that the automation was highly reliable, however, that it might not be perfectly reliable all of the time.

The participants completed one short-duration manual condition of 10 minutes. They then completed both the 10 minute and 30 minute sessions under automation (either information or decision automation), followed by the 10 minute and 30 minute sessions of the other automation type. The order of the tasks was counterbalanced for order effects. Participants filled out the Dundee Stress State Questionnaire (DSSQ) (Matthews et al, 1999) prior to their performing the task. They then filled out post-task DSSQ questionnaires following each condition. The DSSQ provides a multi-dimensional assessment of transient states associated with stress, arousal and fatigue (Matthews, et al, 1999). Participants were also asked to fill out the NASA TLX to measure subjective workload (NASA, 1986), and a trust questionnaire to measure their trust in the automation (Lee and Moray, 1992, 1994).

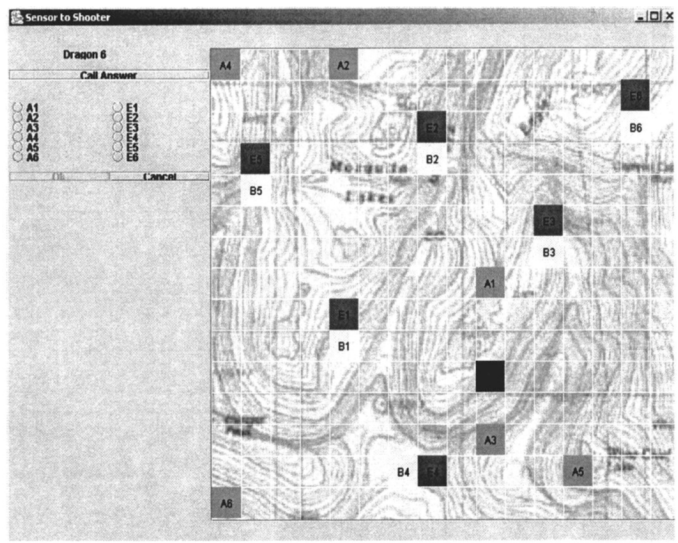


Figure 1: Sensor to Shooter Interface (Manual)

RESULTS

Manual vs. Automation Performance

The mean rate of correct friendly-enemy combat engagement selections was computed for each participant for each condition. A 1-way ANOVA with three levels of a condition factor (Manual, Automated-Reliable Trials, Automated-Unreliable Trials), gave a significant effect for condition, $F(2,18)=51.79$, $p<.05$. Paired-sample t-tests showed that decision accuracy for the reliable automated trials was significantly greater than that on the manual trials, $t(19)=-10.444$, $p<.01$, as well as on the unreliable trials, $t(19)=5.864$, $p<.01$. Figure 2 shows that there was a significant increase in decision-making accuracy with the use of reliable automation support.

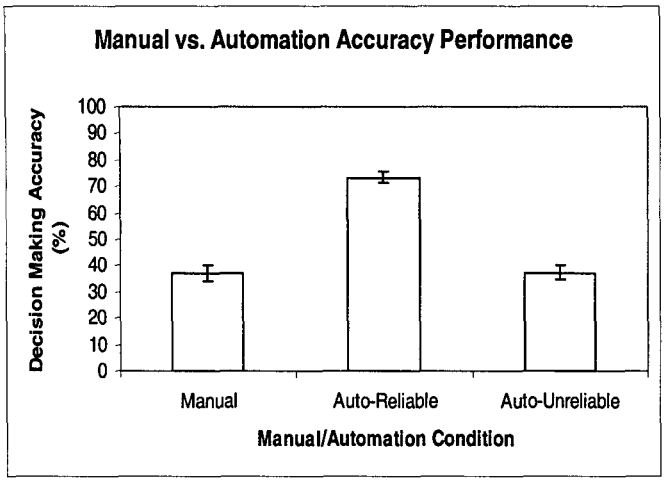


Figure 2: Manual vs. Automation Performance

Next, decision accuracy scores were analyzed with a 2 (Automation Type: Information or Decision) x 2 (Automation Reliability) x 2 (Task Length) ANOVA. The effect of automation reliability was significant, $F(1,19)=34.391$, $p<.01$, reflecting the greater decision accuracy in the reliable compared to the unreliable trials, as shown in the previous analysis. The interaction between Automation Type and Reliability was significant, $F(1,19)=14.586$, $p<.01$. Figure 3 shows that the reduction in decision accuracy for unreliable compared to reliable automated trials was greater for decision automation than for information automation, irrespective of task length.

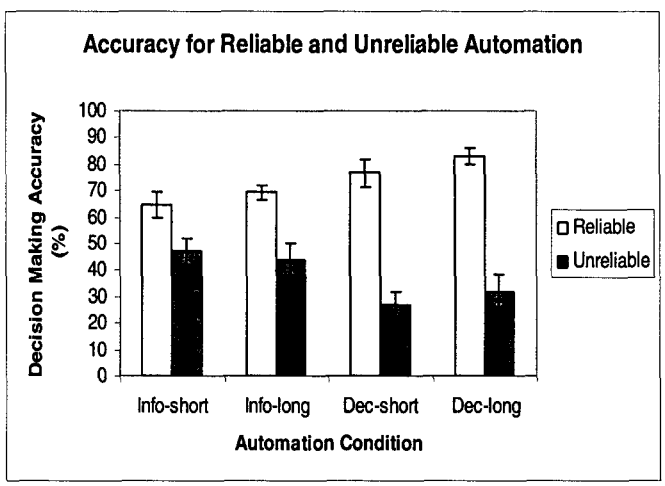


Figure 3: Accuracy for Reliable and Unreliable Automation

Similar analyses of combat engagement selection times showed that response time (RT) for reliable automated trials was significantly shorter than for the manual trials, $t(19)=9.104$, $p<.01$, as well as for the unreliable trials, $t(13)=-7.82$, $p<.01$. There was no significant difference in RTs between manual and unreliable trials, $t(12)=1.635$, $p>.05$. These effects are shown in figure 4. RTs were also analyzed using a 2 (Automation Type) x 2 (Task Length) x 2 (Automation Reliability) ANOVA. The effect of Automation Type was significant, $F(1,13)=22.33$, $p<.01$, with shorter RTs for decision automation.

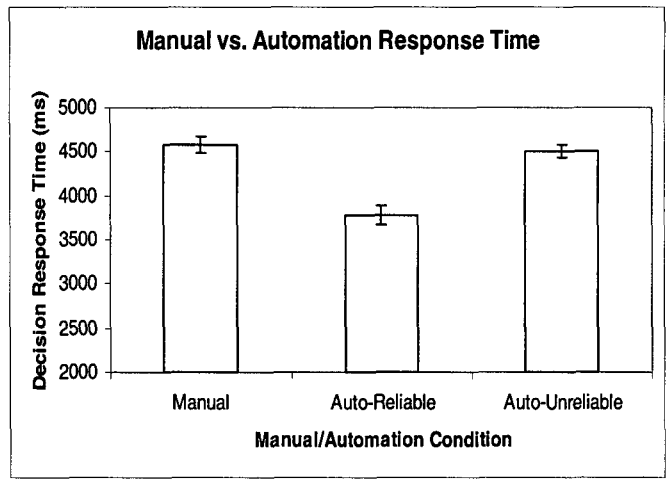


Figure 4: Manual vs. Automation Response Times

The effect of Automation Reliability was also significant, $F(1,13)=61.47$, $p<.01$, reflecting shorter RTs in the reliable compared to unreliable trials. The interaction between Automation Type and Reliability was significant, $F(1,13)=45.487$, $p<.01$. The increase in RT for unreliable trials was greater for decision automation than for information automation. The effect of task length was significant, $F(1,13)=22.545$, $p<.01$, with RTs decreasing for the long task. Finally, the three-way interaction between Automation Type, Task Length and Automation Reliability was significant, $F(1,13)=8.645$, $p<.05$, with RTs on unreliable trials increasing only for decision automation, not for information automation, and primarily for the long task, as shown in figure 5.

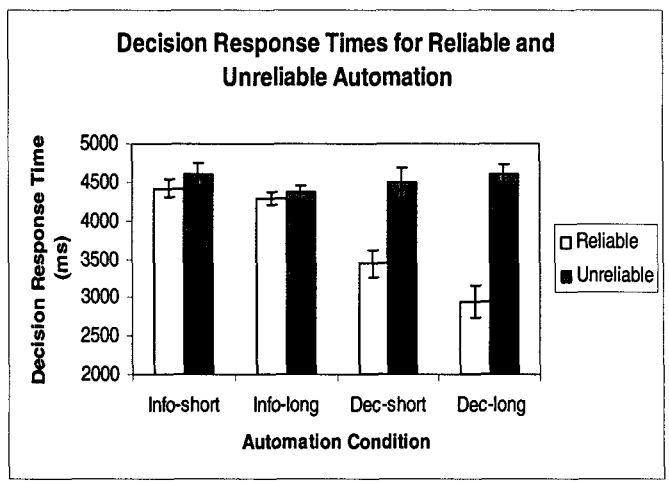


Figure 5: Decision Response Times for Reliable and Unreliable Automation

Stress State Changes

Mean scores on the different aspects of the DSSQ were calculated for each participant pre-task and after each condition. For the analysis, difference scores were used, with each pre-task DSSQ score subtracted from the post-task score. A positive difference score implies that scores on the state measure tend to increase as a consequence of performing the task. A 1-way ANOVA for each component of the DSSQ with 5 conditions (manual, information-short, information-long, decision-short and decision-long) gave significant effects for several of the components: Energetic Arousal, $p<.01$; Motivation, $p<.01$; Self-esteem, $p<.05$; and Task-Related Interference, $p<.01$. Energetic Arousal, Motivation and Task-Related Interference all showed decreases between the pre-task baseline and the automated conditions. The Self-esteem component showed an increase from the manual to the automated conditions. Figure 6 shows the changes in the different stress state components across conditions. Self-focused attention also showed a trend toward a decrease between the pre-task and the automated conditions, $p=.062$.

A 2 (Automation Type) x 2 (Task Length) ANOVA of the DSSQ difference scores gave marginally significant main effects for Task Length on the Task Irrelevant Interference component of the DSSQ, $p<.05$; Anger component, $p=.056$; Concentration component, $p=.067$; and Self-Focused Attention component, $p=.064$. Scores on the Task Irrelevant Component showed that the shorter task had more of a decrease than the longer task, as did the Self-Focused Attention component. The Concentration component showed a larger decrease on the longer task. The Anger component showed an increase on the longer task.

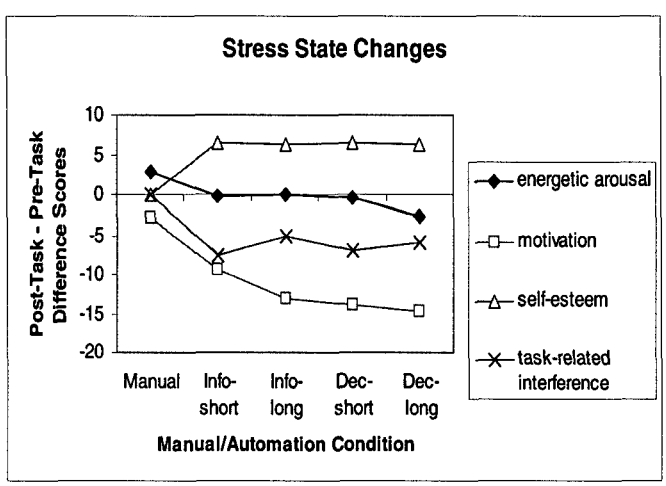


Figure 6: Stress State Changes

Subjective Workload and Trust Ratings

A 1-way ANOVA with five conditions (manual, information-short, information-long, decision-short, decision-long) for each category of the NASA TLX gave significant results only for Temporal Demand, $F(4,15)=3.096$, $p<.05$. Finally, there were no significant changes in subjective ratings of trust between conditions.

DISCUSSION

The present study examined the relative benefits and costs associated with information and decision automation. In addition, we investigated the effects that task length would have on these aspects of performance, as well as on subjective operator stress. The results showed that reliable automation led to a significant benefit in decision accuracy over manual performance. Reliable automation also significantly lowered response times, as compared to unreliable automation and the manual condition. The response times were also significantly reduced on the longer trials, perhaps showing that the participants had developed a greater understanding of the automation when performing the task for longer periods of time. However, there was a cost associated with imperfect automation, such that despite this increased knowledge, participants tended to follow wrong automation recommendations. This tendency was greatly reduced under imperfect information automation. These results are consistent with previous findings (Crocoll & Coury, 1990; Sarter & Schroeder, 2001; Rovira, McGarry & Parasuraman, 2002), but extend them to longer task conditions in which the participant must perform the task for a significantly longer period of time.

The subjective stress levels of participants showed significant changes throughout performance on the task. Energetic arousal showed a significant decrease between the pre-task levels and those of the automated condition, indicative of fatigue. Several of the other state changes found for the DSSQ also indicate that participants found the task to be stressful, especially in the long conditions. The increase in reported anger is indicative of this, as are the decreases in concentration and motivation. In sum, while automation improved accuracy and response times when reliable, the downside was an increase in subjective reports of stress, as well as a performance cost associated with unreliable decision automation.

With respect to the two hypotheses concerning the factors underlying the differential costs of decision versus information automation, the results supported the first and partially supported the second hypothesis. There was clear evidence of differential over-reliance on decision automation which led to poorer performance when wrong recommendations were given. However, there was only partial support for the view that this was due to insufficient development of participant understanding of the automation capabilities, as tested using short and long duration tasks. When greater exposure to automation was provided in the long task, decision times were reduced, but the performance cost of unreliable automation remained, and there was evidence of increased subjective stress and fatigue. The challenge for automation design in future work is to develop procedures that capture the benefit that decision automation provides in shortening overall response times (e.g., reduced sensor-to-shooter time) while minimizing the costs of unreliability. Explicit training, not just greater exposure to automation, may be required. Apart from training, possible mitigating methods include the use of

ecological interface design (EID) to support operator information sampling behavior (Lorenz, Di Nocera, Rottger, & Parasuraman, 2002a), integrated displays (Lorenz, Di Nocera, & Parasuraman, 2002b), and EID displays that explicitly represent automation system intentions and states (Furukawa & Parasuraman, 2003).

ACKNOWLEDGEMENTS

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. We would also like to thank Zihang Ou for his technical support.

REFERENCES

Crocoll, W.M. & Coury, B.G. (1990). Status or recommendation: Selecting the type of information for decision aiding. *Proceedings of the 34th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 1524-1528, Santa Monica, CA: Human Factors Society.

Furukawa, H., & Parasuraman, R. (2003). Supporting system-centered view of operators through ecological interface design: Two experiments on human-centered automation. In *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA: Human Factors Society.

Galster, S. M., Bolia, R. S., Roe. M. M., & Parasuraman, R. (2001). Effects of automated cueing on decision implementation in a visual search task. In *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*, pp. 321-326, Santa Monica, CA: Human Factors Society.

Galster, S., Bolia, R. S., & Parasuraman, R. (2002). Effects of information automation and decision-aiding cueing on action implementation in a visual search task. In *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*, pp. 438-442, Santa Monica, CA: Human Factors Society.

Hancock, P.A. & Mortimer, D. Cleve. (2002). Decision errors under stress. *Unpublished Technical Report*, University of Central Florida. pp. 1-51.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human machine systems. *Ergonomics, 35,* 1243-1270.

Lee, J. & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human Computer Studies, 40,* 153-184.

Lorenz, B., Di Nocera, F., Rottger, S., & Parasuraman, R. (2002a). Automated fault-management in a simulated space flight micro-world. *Aviation, Space, and Environmental Medicine,* 73, 886-897.

Lorenz, B., Di Nocera, F., & Parasuraman, R. (2002b). Display integration enhances information sampling and decision making in automated fault management in a simulated space flight micro-world. In *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomic Society,* pp.31-35, Santa Monica, CA: Human Factors Society.

Matthews, G., Campbell, S. E., & Falconer, S. (2001). Assessment of motivational states in performance environments. In *Proceedings of the Human Factors and Ergonomics Society,* 906-910.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (eds.), *Personality Psychology in Europe* (Vol. 7, pp. 335-350). Tilberg, Netherlands: Tilberg University Press.

NASA Ames Research Center (1986). *NASA Task Load Index (TLX): Paper and Pencil Version.* Moffet Field, CA: NASA Ames Research Center Aerospace Human Factors Research Division.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, & Cybernetics, 30,* 286-297.

Rovira, E., McGarry, K., & Parasuraman, R. (2002a). Effects of unreliable automation on decision making in command and control. In *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*, pp. 428-432, Santa Monica, CA: Human Factors Society.

Rovira, E., Zinni, M., & Parasuraman, R. (2002b). Effects of information and decision automation on multitask performance. In *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*, pp. 327-331, Santa Monica, CA: Human Factors Society.

Sarter, N., & Schroeder, B. K. (2001). Supporting decision-making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors, 43(4),* 573-583.

Temple, J. G., Warm, J. S., Dember, W. M., Jones, K. S., LaGrange, C. M. & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors, 42,* 183-194.

Wickens, C.D., & Xu, X. (2002). Automation trust, reliability and attention. Technical Report AHFD-02-14/MAAD-02-2. Aviation Human Factors Division, Institute of Aviation; Savoy, IL.

FIELD ASSESSMENT OF COGNITIVE PERFORMANCE UNDER STRESS

Wayne C. Harris
University of Minnesota

Peter A. Hancock
University of Central Florida

Performance decrements have frequently been reported in combat but it has been difficult to demonstrate the specific causal source of cognitive degradation sufficient to disrupt performance during military operations. In the present study, seventeen infantrymen were assessed during both a one-week simulated combat exercise and a pre-dawn assault. Significant change did not occur during the one week exercise, however simple reaction time did increase following the assault. One planned field assessment occurred at the moment following a potentially serious medical event. A soldier perceived that he was experiencing heatstroke but medical evaluation indicated that he was still "fit for duty". Cognitive evaluation of his failing state was characterized by slow Simple Response Time and rapid and inaccurate responding on more complex response tasks. Fortunately, the soldier could return to duty but his scores slowly recovered, such that he returned to baseline levels two days later. The present infantry operational environment did not produce significant-information processing change in infantrymen. However, significant performance cognitive decrements followed the specific acute event of assault that was perceived as life-threatening. As we demonstrate, increased reaction time after the assault is consistent with increased fatigue following an extended period in the infantry operating environment.

INTRODUCTION

Systems are being developed which provide soldiers with an impressive array of information and with the promise of increased lethality and survivability through superior situational awareness. Improving soldiers' knowledge of the battlespace may increase their capability, but soldiers must be able to manage those information systems under combat conditions. Performance decrements have been reported in operational settings and simulations that create high stress levels, thus system design should be based on an understanding of the information processing changes that occur during combat (and see Hancock & Szalma, 2003). Performance deteriorates in combat and other high stress activities (Orasanu & Backer, 1996), and in conditions that participants view as life threatening (Berkun, 1964). People ignore training and common sense during moments of great fear and they do not effectively process information during those incidents. The immediate effect of threatening conditions can be a brief, but sometimes catastrophic impairment of decision-making (Morgan, 1995). Decisions are the outcome of the information processing processes, which are composed of a number of sub-tasks or components (Wickens & Flach, 1988). Operational performance capacity has been directly related to these information processing components (see Taylor, O'Hara, Mumenthaler, & Yesavage, 2000). Taylor et al (2000) found that flight simulator performance was linked to visual processing speed, working memory, and math computational skill. Stress disrupts such information processing capabilities (Hancock & Desmond, 2001; Hancock & Warm, 1989; Wickens, 1996) and few settings prove more stressful than combat. However, cognitive task deterioration sufficient to cause the performance deficits associated with combat have been difficult to demonstrate in military settings (Elsmore, Naitoh, & Linnville,1992; Callister, Percival & Retzlaff, 1999; Slaven & Windle, 1999). Infantry are subjected to a variety of conditions that create unique demands and the presence of all or most of those stresses may be required before cognitive deficits appear. Orasanu and Backer, (1996) have divided combat-related stressors into threat and danger, fatigue, workload/information load, and environmental factors such as noise and heat. However, as stress is a function of the individual's reaction to ambient conditions (Lazarus & Folkman, 1984), the effect of these factors on soldiers in the operational environment needs careful assessment. Field training provides a veridical representation of the operational environment with one notable exception, the level of threat and danger is lower in training than in combat. Although danger is reduced, the infantry training environment still remains a demanding one. Assessment of cognitive changes during training can provide an estimate of the relative importance of actual combat stressors to performance deterioration.

A second issue is the timing of the assessment. Combat conditions may produce impaired information processing that produce the reported performance decrements, but soldiers may generate rapid recovery. The result would be that post-combat assessments detect fatigue but not the cognitive changes that change combat performance. The current study developed a stable performance baseline in a low stress setting

and then assessed soldiers in the field while they were participating in combat field exercise. Testing during the exercise eliminated the possibility that the effects had dissipated by the time of testing. Field evaluation also allowed the experimenters themselves direct opportunity to assess the actual stresses present during operation.

METHOD

Participants

Seventeen members of Alpha and Charlie Companies of the Second Battalion, 135th (Air Assault) Infantry participated in this study. Data was collected during a two week infantry field exercise at Ft. Chaffee, AR.

Cognitive Assessment

The ANAM Readiness Evaluation System (ARES) was used to assess cognitive performance. ARES is a selection of Automated Neuropsychologial Assessment Metric (ANAM) tasks presented on a handlheld computer. ANAM (Reeves, Winter, LaCour, Raynsfor, Vogel, & Grissett, 1991) presents a variety of cognitive tasks on a desktop or laptop computer, making adaptation to field research difficult. ARES is a modification of the ANAM for handheld systems. With ARES, it is possible to assess soldiers at their unit's location during breaks in the action. ARES two batteries, Commander and Warrior, were developed to assess fitness for duty. The tasks in the battery here (Chaffee Battery) were expanded to assess components of cognitive performance considered to be important in operation settings (Wickens & Flach, 1988). The battery included Simple Reaction Time (SRT), Logical Reasoning (LRS), Mathematical Processing (MATH), Spatial Processing (Spatial Rotation), Match to Sample (MTS), and a 6-item Sternberg Memory task (MEM). The battery was administered using the handheld Handspring Visor Prism.

Procedures

After reading and signing the informed consent document, volunteers completed the cognitive battery four times in the barracks area. Previous research indicate that performance improves during the first three to four administrations of the battery as participants became familiar with the generalities of the assessment procedures (Benedetto, Harris, & Goernert, 1995). Baseline accuracy and response time were defined as the participant's performance during their final training session and field performance was then compared with this, their last training performance.

RESULTS

Field response times and accuracy were expressed as changes from the soldier's own final baseline session. Data included assessments during day 3 and day 5 of a six day field exercise. The second assessment occurred after the 0500 assault during the second week. An unexpected incident occurred that provided the opportunity to assess a soldier immediately after a high stress incident. This individual's cognitive performance is described in the final section.

Baseline Performance

The performance of most tasks during training was consistent with previous reports (Benedetto, et al., 1995) and the majority of improvement appeared to have been completed by the fourth trial. The first training session error rate was low and error rate remained low throughout pre-exposure. Response time changes during training were expected to improve as a reflection of increased familiarity with the assessment procedure and to be relatively stable by the fourth session. The response time change of Logical Reasoning task in Figure 1 illustrates the typical ARES response time change, which were consistent with the anticipated pattern of change during training. Comparing field performance with a baseline defined as the fourth training trial appears to be a reasonable approach to assessing the effects of field training with the majority of the ARES tasks. One task, Simple Response Time, did not conform to the response time change anticipated during training. Simple Reaction Time increased as training

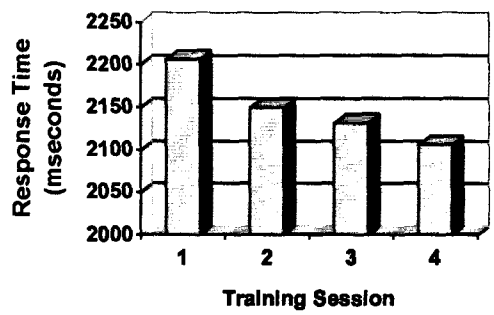


Figure 1. *Logical Reasoning Response time change during training.*

progressed (Figure 2). There was sufficient time between training trials to insure that the increased response time cannot be attributed to fatigue. An alternative possibility is that performance declined after the task was mastered. Simple

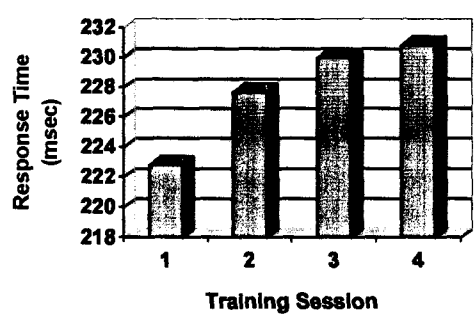


Figure 2. *Simple Response Time change during training.*

Reaction Time was the first task in the battery. After the first trial, soldiers knew that they would be completing a series of challenging tasks and they may have exerted less effort on the

apparently simple first task. The atypical changes during Simple Response Time training affect the use of the fourth training trial as a performance baseline and this an issue that will be discussed in the interpretation of the field data.

The two-week field training was composed of two exercises. The first exercise was a six-day operation in which the battalion's mission was to slow an invading force. The invading force's attack began along a road in the sector defended by Alpha Company. In addition to the differing levels of engagement, Alpha Company's sector was a rocky, densely wooded area and Charlie Company was on a ridge protecting a pass which was cooled by a breeze. Analysis of the field data therefore begins with an analysis of the entire sample, but a secondary analysis is performed on the company operating under the more difficult conditions. The second exercise was an 0500 assault. The operating conditions of the two companies did not differ during assault and the single analysis included both companies.

Week One Field Operations

Since the operating conditions of Alpha Company were more demanding than those of Charlie Company, two analyses were performed. The first analysis included data from all soldiers and the second analyses was limited to the data from Alpha Company alone. The low number of soldiers in the second sample limited the power of the analysis and significance levels of approximately 0.10 to 0.19 are reported as trends that warrant further investigation.

Week One Field Operations Day 3

Accuracy. No accuracy change was found in the analysis of the entire sample or in the analysis of Alpha Company, the company which operated in the more difficult setting. The only change that was of sufficient magnitude to warrant further examination was decreased Alpha Company Spatial Processing accuracy ($p = .16$).

Response Time. Response time did not differ from the final baseline value with one exception. Simple Response Time was faster than baseline response time during the first day of field training. On Field Day 3, the entire sample Simple Response Time decreased 13 milliseconds (ms) from the last baseline trial ($p = .02$) and mean Alpha company response time decreased 16 ms ($p = .01$). Interpretation of these finding requires reference to performance changes during training. Simple Response Time increased during the training period, thus the apparent improvement in Simple Response Time reflects a return to the levels observed when the task was first presented.

Week One Field Operations Day 5

Accuracy. Error rate of the combined sample of Alpha and Charlie Companies did not differ from baseline assessment error rate and no scale change approached significance. Examination of the accuracy data from Alpha Company indicated that increased Spatial Processing error rate continued to be greater than baseline on Field Operation Day 5 ($p = .11$), suggesting that Spatial Processing changes warrant further examination.

Response Time. The Simple Reaction Time of Alpha Company remained significantly faster than their baseline on Operational Day 5 ($p = .04$). A significant response time change on the complex cognitive tasks was not found when the data from Alpha and Charlie Companies was analyzed. There was also a trend toward faster responding to the Spatial Processing task. Alpha Company Spatial Processing reaction time decreased and the change approached significance ($p = .10$). Thus Spatial Processing became faster although error rate was still above baseline level.

Week One Field Operations Summary

Field Assessment during the five days of field exercises provided little indication that the operating conditions affected soldier's cognitive performance. The continuous presence of marginal Spatial Processing decrements suggest that spatial tasks should be examined in greater detail in future studies. Rather than interpreting Simple Response Time changes as improved capacity, the Simple Reaction Time improvement appears to be related to changes during training that produced an inaccurate estimate of baseline performance. Soldiers mastered the Simple Reaction Time task before the remaining tasks were learned. The initial trial apparently provided a better estimate of performance capability than later trials. When presented as a performance assessment in the field, soldiers apparently increased effort to their initial level.

Week Two Assault

Low stress baseline performance level was re-established on Sunday of training week two. On Tuesday soldiers prepared for an assault the following morning and that evening they traveled by truck to a forward position. Troops began their movement toward the enemy on foot at approximately 0200 on Wednesday and the Battalion attacked at an 0500. The assault was completed at approximately 0900 and a cognitive assessment was obtained while troops waited for transportation buck to the barracks area.

Accuracy. Pre-dawn assault accuracy did not differ from accuracy obtained in the barracks three days earlier.

Response Time. In contrast to earlier field observations, post-assault Simple Response Time was longer than baseline assessments ($p = .05$). These changes are similar to changes found after a week of stressful survival evaluate (SERE) training (Hancock & Harris, 2000) and suggest that the soldiers' fatigue level had increased after two weeks of intense training and an approximately 26 hour operation. There was no evidence of complex cognitive task impairment after the assault. A trend toward slower responses to the Memory task ($p = .16$) suggests that memory changes should also be examined in future studies which the effect of extended field operations and this is now underway.

High Stress Individual

The demands on individual soldiers within each company and their ability to manage those demands varied. On day three of field training, a number of soldiers were experiencing difficulty adjusting to the heat and high workload. When Alpha Company was contacted in the field, two soldiers were experiencing symptoms suggesting acute heat stress. Medics evaluated the soldiers and one soldier was evacuated from the field however the second soldier was returned to duty. Since he had been evaluated and found fit for duty, the second soldier was still available for testing. He exhibited no obvious signs of impairment and he completed the field exercise without further problems. However, the results of the cognitive assessment indicate the presence of significant deterioration of the complex tasks in the ARES battery immediately following the incident. Baseline performance of the soldier could be characterized as slightly faster and less accurate than the platoon average (Table 1), and his performance change from baseline immediately following the incident does not appear be an artifact of an inappropriately high baseline level.

The Simple Reaction Time of the high stress soldier increased 25 ms. on Field Day 3. This change contrasted with the Alpha Company mean which had decreased 15.83 ms. from their last baseline session on Field Day 3. Alpha Company baseline mean error rate was less than 2 and changes from baseline during the first field assessment were not significant. In contrast, the high stress soldier's error rate increased dramatically. The change from baseline for Alpha Company and the high stress soldier are presented in Table 1.

Table 1: *Mean number of errors of Alpha Company and high stress soldier during baseline assessment and Field Day 3*

	Alpha Company		Soldier	
	Baseline	Day 3	Baseline	Day 3
Logical Reasoning	1.82	1.92	4	11
Mathematical Processing	2.45	1.25	2	11
Spatial Processing	1.18	2.08	5	2
Spatial Memory	1.18	1.46	2	8
6-item Memory	2	1.23	8	21

Complex task reaction times of Alpha Company did not change from their final baseline level, however, the high stress soldier's response times decreased to 200-300 ms, a time typical for the Simple Reaction Time task. Response times of 1000-2000 ms. are normally found on complex task performance. The soldier's rapid responses suggest that he was not taking sufficient time to complete the cognitive operations needed to complete the task for most items. There was no evidence that the individual had taken a strategy such as pressing one key in response to all items.

Table 2: *Mean response time of Alpha Company and high stress soldier during baseline assessment and Field Day 3*

	Alpha Company		Soldier	
	Baseline	Day 3	Baseline	Day 3
Simple Response Time	231	202	195	220
Logical Reasoning	2083	1788	2085	175
Mathematical Processing	1841	2051	1665	210
Spatial Processing	1881	2082	1130	50
Spatial Memory	1210	1305	1040	295
6-item Memory	660	685	740	200

DISCUSSION

Performance deterioration during combat (Marshall, 1947) and simulated combat (Berkun, 1964) indicates soldier vulnerablility and the addition of more cognitively demanding technological is likely to increase the severity of such combat performance deterioration (Orasanu & Backer, 1996). Operational performance is related to the individual's ability to perform cognitive operations (Taylor et al., 2000), however, assessment of military personnel's cognitive performance during apparently stressful periods have often not detected significant information processing changes. The current study assessed cognitive performance during a field exercise that included conditions present during combat with the exception of threat of enemy fire. Consistent with previous reports, physical discomfort, disrupted sleep, potential danger and other apparent stressors in the infantry operating environment appear to have minimal effect on soldier's cognitive ability. Accuracy and response time were stable during 5 days of field operations. The small number of soldiers assessed may have been insufficient to detect small changes, but there was no evidence of cognitive changes sufficient to explain reported decrements in combat settings. Trends were noted that suggest increasing spatial processing errors that could be the focus of future studies. Increased Simple Reaction Time after the pre-dawn assault is consistent with other reports of the effect of extended stress (Hancock & Harris, 2000). Extended periods in the infantry environment resulted in a condition more appropriately described as chronic fatigue than distress per se.

The fortuitous presence of the researcher at the moment when an individual was experiencing an event that was an apparently serious medical problem provided the opportunity to measure cognitive changes during a period of high threat. Although caution must be exercised when interpreting the data from one individual, the large cognitive task performance decrements observed are consistent with observations of performance deterioration in combat. Rapid, inaccurate performance was observed immediately following the

potentially dangerous medical situation. The soldier however did not manifest gross symptoms of impairment. He did not indicate that he could not perform the assessment task, and he returned to duty and on the surface performed adequately for the remainder of the training. By two days later, his cognitive performance had returned to baseline levels. The results suggest that during and immediately after periods of high stress, soldiers' cognitive capacity deteriorates significantly but such cognitive decrement might not be immediately apparent to his colleagues and commanders. Although it cannot be assumed that performance on essential military tasks would exhibit an equal decrement, the data do indicate that high threat situations produce brief periods of impaired decision making. In addition, data suggest that impairment can be detected by brief, portable cognitive testing systems that may in future be embedded into the soldier's actual mission-related tasks. The rapid onset of stress followed by a precipitous deterioration of performance is consistent with reports of grossly impaired decisions during stress (Berkun, 1964; Morgan, 1995) further field evaluations are in progress to clarify such complex relations.

The results are also consistent with theoretical models of stress that predict minimal performance disruption while an individual is capable of adapting to stressors, and a precipitous deterioration when demands exceed their coping capacity (Hancock & Warm, 1989). The Hancock and Warm model indicates that as conditions become more demanding, performance will be relatively stable until demands exceed the person's threshold capacity to adapt. Only after demands exceed resources does performance begin to deteriorate. The presence of generic stressors cannot be assumed to cause stress. Soldiers show the ability to cope with a variety of conditions that few would argue are pleasant, but soldier's reaction to those conditions suggest they are not experiencing stress directly. The comment when training conditions were discussed was "This is where we work". Of the combat-stressors threat, workload, environmental conditions and fatigue (Orasanu & Backer, 1996) danger appears to be the factor most related to catastrophic performance deterioration. The non-linear relationship between stress and performance deterioration makes it dangerous to extrapolate the findings of studies using low stress levels to predict effects in combat.

The post-assault assessment provides the strongest evidence of a cognitive performance change. While performance on the complex tasks was maintained, SRT increased following the assault. The result is consistent with previous reports that Simple Reaction Time increases following extended periods of stress (Hancock & Harris, 2000). Although complex task performance is sensitive to stress (Broadbent, 1971), increased effort can improve complex task performance more than it can simple task performance (Heuer, Spijkers, Kiesswetter, & Schmidtke, 1998). Complex cognitive tasks deficits can therefore be masked during brief assessments. The results support the position that although the infantry environment produces compensable stress, operational conditions increase fatigue and the assessment we have found most sensitive to fatigue is Simple Reaction Time.

ACKNOWLEDGEMENT

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, P.A. Hancock, Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

REFERENCES

Benedetto, J., Harris, W. C., & Goernert, P. N. (1995). Assessing gender differences and norm data on a cognitive performance measure. *Proceedings of the Human factors and Ergonomics Society 39th Annual Meeting*, San Diego, CA, October, 1995.

Berkun, M. M. (1964). Performance decrement under psychological stress, *Human Factors*, **6**, 21-30.

Broadbent, D.E. (1971). *Decision and stress*. London: Academic Press.

Callister, J. D., Percival, G. L., & Retzlaff, P. D. (1999). Stress and fatigue during aircrew survival training. *Proceedings of the Tenth International Symposium on Aviation Psychology*, **10** *(2)*, 843-847.

Elsmore, T. F., Naitoh, P., & Linnville, S. (1992). *Performance assessment in sustained operations*. Naval Health Research Center Report No. 92-30

Hancock, P.A. & Desmond, P.A. (2001) (Eds.). *Stress, workload and fatigue*. Erlbaum: Mahwah, NJ.

Hancock, P. A. & Harris, W. C. (2000). *Information processing following extended stress*. Final Report on Office of Naval Research Grant # N00014-99-1-1034.

Hancock, P.A., & Szalma, J.L. (2003). Operator stress and display design. *Ergonomics in Design*, **11**, (2), 13-18.

Hancock, P. A. & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, **31**, 519-537.

Heuer, H., Spijkers, W., Kiesswetter, E. & Schmidtke, V. (1998). Effects of sleep loss, time of day, and extended mental work on implicit and explicit learning. *Journal of Experimental Psychology: Applied*, **4**, 139-162.

Lazarus, R., & Folkman,S. (1984). *Stress, appraisal, and coping*. New York: Springer Verlag.

Morgan, W. P. (1995). Anxiety and panic in scuba divers, *Sports Medicine*, **20***(6)*, 398-421.

Orasanu, J. M., & Backer, P. (1996). Stress and military performance. In: J. Driskell & E. Salas (Eds.), *Performance under stress*. (pp. 89-117). Hillsdale, NJ: Lawrence Erlbaum Associates.

Reeves, D.L., Winter, K.P., LaCour, S. J., Raynsfor, M., Vogel, K., & Grissett, J. D. *(1991). The UTC-PAB/AGARD stress battery: User's manual and system documentation.* NAMRL Special Report 91-3.

Slaven, G. M. & Windle, C. M. (1999). Cognitive performance over 7 days in a distressed submarine. *Aviation, Space, and Environmental Medicine*, **70***(6)*, 604-608.

Taylor, J. L., O'Hara, R., Mumenthaler, M. S., & Yesavage, J. A. (2000). Relationship of CogScreen-AE to flight simulator performance and pilot age, *Aviation, Space, and Environmental Medicine*, **71**, 373-380.

Wickens, C. D. (1996). Designing for stress. In J.E. Driskell & E. Salas, (Eds), *Stress and human performance*, Mahwah, NJ: Lawrence Erlbaum.

Wickens, C. D. & Flach, J. (1988). Human information processing. In E. Wiener & Nagel (Eds.), *Human factors in aviation*. (pp 111-155), New York: Academic Press.

THE EFFECTS OF TRUST, COMMUNICATION AND PROXIMITY ON TEAM PROBLEM SOLVING

Nicole Pruitt, Jessica Bolton, Andy Kearney, Benjamin Swan
Kansas State University

Kip Smith
Linköping University/Institute of Technology

This paper discusses an experiment on the interaction of trust, communication , and proximity on team problem solving. The research is motivated by the US Army's plan to change the organizational structure of the combat infantry from a close-knit brotherhood to an alliance of high-tech autonomous individuals. We are interested in ascertaining the possible consequences this transition may have on teamwork in a context that generalizes to the Army's mission. In our experiment, communication and proximity were fully crossed in a 2X2 design. Trust at three levels – trust, distrust, and betrayal – was a subject variable revealed by play of an economic game. We find a significant main effect for mode of communication: teams that communicate freely work more effectively. Further, we find a significant main effect for trust. When team members trust each other, they work more efficiently. The lack of a significant interaction between proximity and communication reveals that the quality of team problem solving is not sensitive to the mode of communication. Teams that coordinated their activity using a radio performed no worse than teams that worked face-to-face. This finding bodes well for organizations that rely on remote command and control.

INTRODUCTION

There is a long tradition of research on the issue of communication and trust and how they interact to affect individual and team performance (e.g., Latané, 1981; Milgram, 1974). This paper contributes to that tradition by testing the impact of communication, trust, and proximity on two-person team performance in a series of laboratory tasks.

The research discussed here was triggered by the development of the US Army's Objective Force Warrior, more familiarly known as 'the Army of One.' The Army is changing the organizational structure of the combat strike unit from larger, more costly troops to a one-person format in which the soldier receives command and control by means of radio transmission. We are interested in ascertaining the possible consequences this transition may have on teamwork in a context that generalizes to the Army's mission. Does the transformation from a close-knit brotherhood to an alliance of high-tech autonomous individuals have consequences on mission success? In particular, what is the impact of the change in mode of communication from face-to-face to radio transmission on team efficiency and effectiveness?

In previous eras of warfare the bond formed by soldiers in their unit was a motivator. The face-to-face communication aided assurance of the completion of orders. With the transformation to technology-driven battlefields, a smaller unit of soldiers will be given the responsibility to carry out a mission previously tasked to a larger unit. These soldiers will have to rely and place their trust in technology rather than in face-to-face contact with each other and their commanders.

Our research draws upon social impact theory (Latané, 1981) to explain the interaction between communication and trust on team performance. This theory suggests that the presence (or absence) of an audience, or co-participant, influences individual and team behavior. A leader who is on-site and communicating directly with subordinates has a greater impact on their behavior than a leader who is distant or communicating electronically (Latané & L'Herrou, 1996). Empirical work supporting social impact theory has found that the possibility of being evaluated by co-participants may also alter behavior (Latané & Wolf, 1981). Social impact theory leads us to expect

performance will be better when decision makers are colocated rather than dispersed.

We expect our research to show that three factors influence team performance: communication mode, degree of proximity, and category of trust (trust, distrust, or betrayal). First, we expect performance in a distributed problem-solving task to improve if participants are permitted to communicate freely. Free communication can be either direct or mediated. Free talk is direct when spoken face-to-face. Mediated free talk occurs over electronic links. We predict that team performance will be better under both free-talk conditions than in the conditions in which when participants are not allowed to communicate.

In addition, we expect trust to influence team performance. Trust necessarily involves an interaction between two or more individuals or groups (Costa, Roe, & Taillieu, 2001). For trust to exist the party that acts first (or whose action is revealed first) must exhibit trusting behavior and the party that responds must prove to be trustworthy. Following Rousseau et al. (1998), we define trusting behavior as the willingness to be vulnerable to the actions of others (the second party, those who are trusted) based upon positive expectations of the others' behavior. As such, trust is more than cooperation; it is cooperation given a palpable likelihood of (opportunity) loss - risk - and an expectation of positive reciprocation (Lewicki, McAllister, & Bies, 1998; Yates & Stone, 1992). Trustworthiness supports behavior that honors the first party's implicit expectation of positive reciprocation. It often, but not necessarily, incurs some measure of opportunity loss.

A lack of trust can be manifested in two distinct ways, by either distrust or betrayal. Distrust is marked by the absence of trusting behavior by the party to act first while betrayal is marked by the absence of trustworthy behavior by the party that responds (Burnham, McCabe, & Smith, 2000). If our participants are both trusting and trustworthy, they are more likely to perform well as a team than if they distrust each other or experience betrayal.

In addition to studying the interaction of communication, proximity and trust on team performance, we are interested in their impact on individual cognitive performance. We have no *a priori* reason to expect that these factors influence individual performance. In particular, we expect to observe no detrimental effect of communication mode or trust condition on individual cognitive performance. Accordingly, we expect to find a dissociation between our measures of team and individual performance.

METHOD

Participants

To test these hypotheses, we asked 72 pairs of participants (18 female pairs, 23 male pairs, and 31 mixed pairs; mean age 19, mode 18, maximum 48) to engage in three games across a computer network. Participants were drawn from a population of general psychology students who received class credit for participation. All participants signed informed consent forms and were treated according to APA guidelines.

Design

The experiment manipulated proximity and communication in a fully crossed 2 X 2 design with four true experimental levels: colocated with free talk, colocated with no talk, remote with free talk, and remote with no talk. In the colocated conditions, participants sat side-by-side in the same room. In the remote conditions, they sat in separate rooms. In the no-talk treatments they were not allowed to communicate at all. In the colocated free-talk condition, they are allowed to communicate freely. In the remote free-talk condition, they communicated using two-way radios.

As part of the experiment, pairs of participants in all four conditions made choices that revealed levels of trust. Their choices self-selected the pair into one of three categories: trust, distrust, or betrayal. This quasi-variable serves as the third factor in our design.

Procedure

Pairs of participants were randomly assigned to one of four conditions defined by two levels of communication (free and none) and two levels of proximity (colocated and remote) and asked to play three computer games. The first was a test of individual cognitive processing, the mental rotation task of Shepard and Metzler (1971). The second task was the first of the two-person team tasks, the extensive form economic trust game of McCabe, Rassenti, and Smith (1996), shown in Figure 1. The teams then played a variant of the classic Hobbits and Orcs river-crossing problem (Greeno, 1974). Finally, each participant individually performed a second trial of the mental rotation task.

The mental rotation task, an analog coding approach to studying perception of mental imagery, presented two three-dimensional objects on a computer screen. Participants were asked to identify whether the objects were the same or different (mirror images). To respond, the participant pressed the 'Y' key to answer "Yes" to the question "Are they the same?" and the 'N' key to answer "No." The dependent variables were reaction

time and accuracy. This task was administered both before and after the team tasks in a within-subjects (individual) repeated-play design. Learning should allow the participants to improve across repetitions. Any decrement would suggest that the intervening team tasks have a detrimental effect on individual cognitive performance. The software platform for the mental rotation task was SuperLab Pro (© Cedrus Corporation).

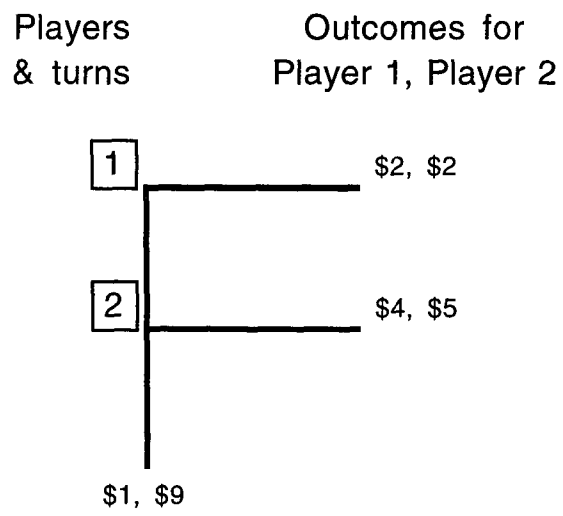


Figure 1 - The extensive-form trust game used as the stimulus in the quasi-manipulation of trust. The numbers in the squares represent Player 1 and Player 2. Player 1 has the first move. A move to the right is a distrusting move that nets both players $2. A move down is a trusting move that passes play to Player 2. A move right by Player 2 is a trustworthy move that nets Player 1 $4 and Player 2 $5. A move down by Player 2 betrays Player 1's trust by awarding $9 to Player 2 but only $1 to Player 1.

The first team game was the trust game shown in Figure 1. Participants played for US dollars that were received immediately upon completing the game. In this game, participants randomly assigned the role of player 1 make a decision to move right or to move down. Moving to the right ends the game, gives both players $2, and places the team in the distrust condition. Moving down passes play to player 2 and is an explicit statement of trust. By moving down, player 1 is trusting player 2 to move to the right, netting both a larger payoff. If Player 2 is given the chance to play, he or she may choose to move right or down. Moving right nets player 1 $4 and player 2 $5, ends the game, and places the team in the trust (and trustworthy) condition. Moving down is an explicit betrayal of player 1's trust. This move nets player 2 $9 while player 1 receives only $1. It also ends the game and places the team in the betrayal condition. Each team played the game only once and, in the process, placed itself into one of three categories of trust – trust (and trustworthy), distrust, and betrayal. This self-selection is a participant variable and the resulting treatments are quasi-experimental variables. The design is accordingly 2 x 2 x 3 (communication by proximity by trust). We did not collect subjective ratings of trust to avoid inducing response bias.

The second team task was a variant of the Hobbits and Orcs river-crossing problem. The game starts with three Hobbits and three Orcs on one side of a river. They want to cross the river using a small boat. The goal of game is to move them all safely across the river. Three rules constrain play: (1) the Orcs may not outnumber the Hobbits anywhere at any time (or else the Orcs eat the Hobbits), (2) a character must be in the boat to get it across the river, and (3) no more than two characters may be in the boat at one time. In our variant of the game, one player is responsible for loading and launching the boat from the left bank, the other player for loading and launching the boat from the right bank. Players are given 100 moves or 45 minutes to complete the game. Play in the Hobbits and Orcs game gives us three measures of team problem solving: the amount of time taken, the number of moves, and the number of illegal moves (e.g., feeding a Hobbit to the Orcs). The platform for both the trust game and the Hobbits and Orcs river-crossing problem was Java script written for this experiment.

Analysis

The team is the unit of analysis except for the mental rotation task in which the individual is the unit of analysis. We used a within subject t-test to analyze the mental rotation data and 2 x 2 x 3 ANOVA to analyze the dependent measures from the Hobbits and Orcs game.

RESULTS

By manipulating communication and proximity, and by allowing participants to reveal trusting behavior at three levels, and by measuring performance in the Hobbits and Orcs game, we test whether communication, proximity, and trust interact to influence team performance in a cooperative problem-solving task. A box plot analysis of the performance data from the 72 teams was conducted. Seven outliers were revealed and eliminated from all analyses.

Figure 2 shows the time on task data aggregated across the four conditions of communication and proximity to show the main effect of trust on performance. The ANOVA reveals that trust is significant $F(2, 53) = 4.52$, $p < .016$. Teams that were

trusting (and trustworthy) successfully completed the task in less time. No other main effects or interactions were significant for time on task.

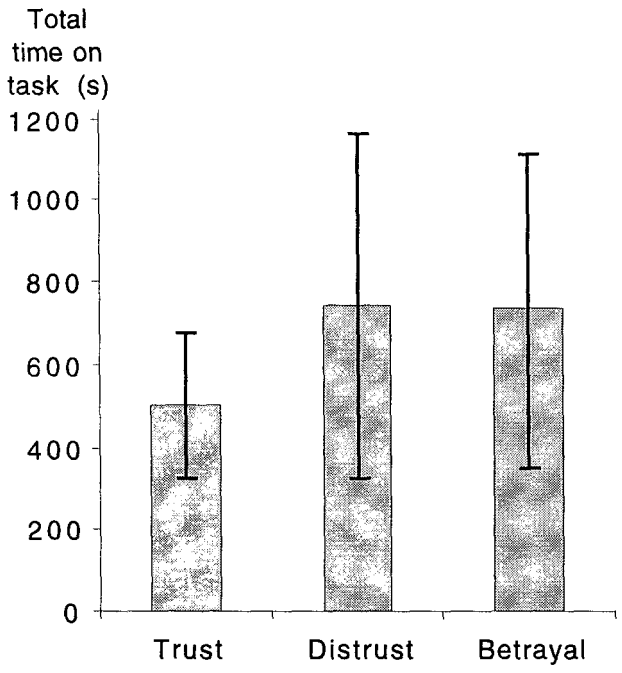


Figure 2 – Graph showing the statistically significant main effect of trust on the time for task completion. Errors bars are one standard deviation. Trusting teams took less time than teams that distrusted or betrayed.

Figure 3 shows the data for the total number of moves aggregated across proximity and trust to show the main effect of communication on performance. The ANOVA reveals that communication is significant $F(1, 53) = 8.69$, $p < .005$. Teams that conversed freely (either face-to-face or over the radio) completed the task more effectively. The three-way interaction is also significant $F(2, 53) = 3.77$, $p < .030$. Teams in the two no-talk conditions who trusted and teams in the remote no-talk condition who experienced distrust needed more moves. Teams that experienced betrayal were equally (in)efficient across all conditions.

Figure 4 shows the data for the number of illegal moves showing the main effect of communication on errors made. The ANOVA reveals that communication is significant $F(1, 53) = 10.79$, $p < .002$. Teams that communicated freely made fewer errors. No other main effects or interactions were significant for illegal moves (errors made).

Of the 130 participants, 125 were both faster and more accurate in the second trial of the mental rotation task. This improvement reflects learning. As expected, the manipulations of communication, trust, and

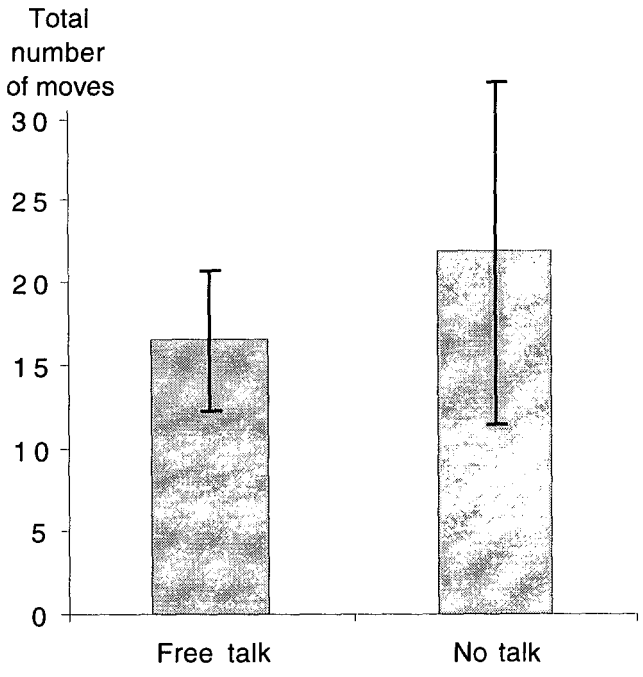


Figure 3 – Graph showing the statistically significant main effect of communication on total number of moves taken to complete the game. Errors bars are one standard deviation. Teams that were allowed to talk made fewer errors.

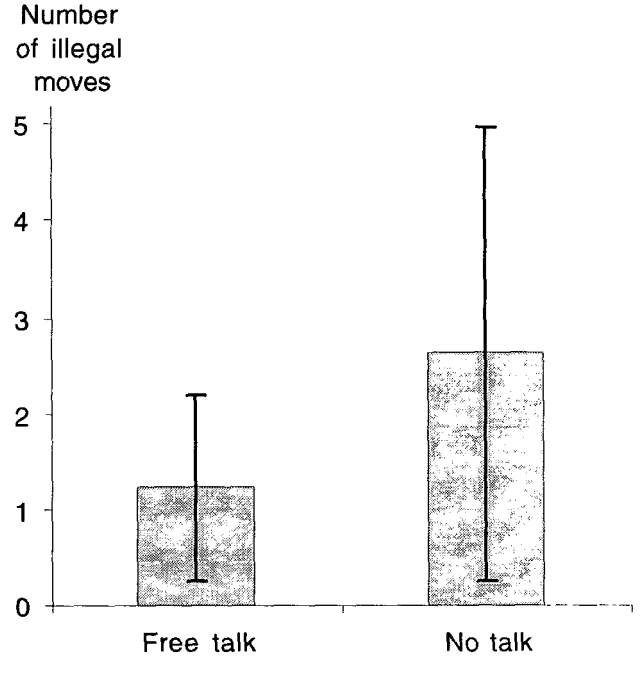


Figure 4 – Graph showing the statistically significant main effect of communication on errors. Errors bars are one standard deviation. Teams that were allowed to talk made fewer errors.

proximity had no effect on this individual cognitive process. Accordingly, we have documented a dissociation of the effect of trust, communication on, and proximity on team and individual performance.

DISCUSSION

The performance data indicate that trusting teams are faster at completing the task. Further, teams that communicate complete the task more effectively and make fewer errors. It comes as no surprise that open communication improves performance. However, the effect of trust on team efficiency is relatively unexpected. Both colocated and distributed teams completed the problem-solving task more efficiently if they previously experienced trust. Distrust and betrayal are detrimental to team efficiency.

In contrast, our data show no main effect for proximity. Teams that were located in different rooms were both as efficient and as effective as teams that worked face-to-face. This bodes well for the design of distributed work groups. For distributed groups to work well, however, they must trust each other and communicate freely.

The long-term goal of this study is to identify any deleterious effect on cognitive performance that might be induced or exacerbated by the erosion of trust that may be induced by the US Army's planned transition from cohesive fighting units to ad-hoc assemblies of specialists. This study suggests that team problem solving suffers in the absence of trust. It also suggests that the extensive form trust game is an effective experimental tool for eliciting trust, distrust, and betrayal. We plan to continue to use the trust game to identify situations in which cognitive performance suffers in the absence of trust. We also plan to conduct experiments that introduce manipulations that are expected to mitigate those effects.

ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Dr. Elmar Schmeisser is the Technical Monitor. The Java scripts for the trust game and the Hobbits and Orcs river-crossing problem were written by Blaine Williams.

REFERENCES

Burnham, T., McCabe, K., and Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. Journal of Economic Behavior and Organization **43**, 57-73.

Costa, A. C., Roe, R. A., & Taillieu, T. (2001). Trust within teams: The relation with performance effectiveness. European Journal of Work and Organizational Psychology **10** (3), 225-244.

Greeno, J. G. (1974). Hobbits and Orcs: Acquisition of a sequential concept. Cognitive Psychology **6** (2), 270-292.

Latané, B. (1981). The psychology of social impact. American Psychologist **36**, 343-356.

Latané, B., & L'Herrou, T. (1996). Spatial clustering in the conformity game: Dynamic social impact in electronic groups. Journal of Personality and Social Psychology **70** (6), 1218-1230.

Latané, B., & Wolf, S. (1981). The social impact of majorities and minorities. Psychological Review **88** (5), 438-453.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. Academy of Management Review **23** (3), 438-458.

McCabe, K. A., Rassenti, S. J., & Smith, V. L. (1996). Game theory and reciprocity in some extensive form experimental games. Proceedings of the National Academy of Sciences, USA **93**, 13421-13428.

Milgram, S. (1974). Obedience to Authority. New York: Harper & Row.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Academy of Management Review **23** (3), 393-404.

Shepard, R. N. & Metzler, J. (1971). Science 171 (3972), 701-703.

Yates, J. F., & Stone, E. R. (1992). The risk construct. In J. F. Yates (Ed.) Risk Taking Behavior. New York: Wiley.

COMMUNICATION, PROXIMITY, AND TRUST IN DISTRIBUTED COMMAND AND CONTROL

Benjamin Swan, Beth Younger, Jana Smith, Molly Toll
Kansas State University

Kip Smith
Linköping University/Institute of Technology

This paper discusses an empirical study of the effects of communication mode and trust on performance in a distributed command and control task. The study was motivated by the US Army's plan to transition from conventional ground forces to the Objective Force Warrior, an army of autonomous, dispersed individuals. We begin with a review of this change in organizational structure in the light of social impact theory. We then develop hypotheses relating trust, proximity, and modes of communication to team performance and discuss our laboratory experiment. Trust interacts significantly with both communication and proximity. The failure to reciprocate trust appears to make it more difficult for teams to communicate and to work face-to-face. If this result generalizes, it may have profound implications for the design of distributed work groups.

INTRODUCTION

Combat soldiers have always been taught, from day one in boot camp, the importance of camaraderie. They train together, endure hardships together, learn together, and grow together. They maintain close contact, both in terms of physical proximity and direct communication. In the process, they form the bonds of friendship, team identity, and trust. Their effectiveness as a team has been rooted in their open communication, close proximity, and mutual trust. The US Army is now flying in the face of this tradition. The cohesive team will be replaced by the Objective Force Warrior, otherwise known as the 'Army of One.' Camaraderie, communication, and trust will no longer be emphasized. Teams will no longer have face-to-face contact with their superior officers, but will get their orders via radio links from someone who may be miles or a continent away. It is unknown how this change in organizational structure may effect team decision making and performance.

Previous research has supported the intuition that the openness of communication between two or more people influences the level of trust between those people (Latané & Wolf, 1981; Latané & L'Herrou, 1996). Empirical work supporting social impact theory has found that people who communicate freely tend to trust each another (Lewicki & Bunker, 1996). In contrast, people who have never met or seen each other are more likely to distrust or to betray. It stands to reason that the nature of trusting behavior (e.g., trust, distrust, or betrayal) will influence the degree of cooperation between team members performing a distributed command and control task. The research discussed here tests that hypothesis.

We expect our data to show that three factors influence team performance: communication mode, proximity, and trust (trust, distrust, or betrayal). First, we expect performance in the distributed command and control task to improve if participants are permitted to communicate freely, either face-to-face or over electronic links like radios. We predict that team performance will be better under these conditions than when participants perform their portion of the distributed task anonymously. We also expect performance to be better if our participants are collated rather than dispersed. When participants can see each other, we predict better performance.

The final factor that we expect to influence team performance is trust. Trust necessarily involves an interaction between two or more individuals or groups (e.g., Costa, Roe, & Taillieu, 2001). For trust to exist the party that acts first (or whose action is revealed first) must exhibit trusting behavior and the party that responds must prove to be trustworthy. Following Rousseau et al. (1998), we define trusting behavior as the willingness to be vulnerable to the actions of others (the second party, those who are trusted) based upon positive expectations of the others' behavior. As such, trust is more than cooperation. It is cooperation given a palpable likelihood of (opportunity) loss - risk - and an expectation of positive reciprocation (Lewicki, McAllister, & Bies, 1998; Yates & Stone, 1992). Trustworthiness supports behavior that honors the first party's implicit expectation of positive reciprocation. It often, but not necessarily, incurs some measure of opportunity loss. A lack of trust can be manifested in

two distinct ways, by either distrust or betrayal. Distrust is marked by the absence of trusting behavior by the party to act first while the latter is marked by the absence of trustworthy behavior by the party that responds. If our participants trust their counterparts, they are more likely to perform well as a team than if they distrust each other or experience betrayal.

In addition to studying the interaction of communication, proximity, and trust on team performance, we are interested in their impact on individual cognitive performance. We have no *a priori* reason to expect that these factors influence individual performance. In particular, we expect to observe no detrimental effect of communication mode or trust condition on individual cognitive performance. Accordingly, we expect to find a dissociation between our measures of team and individual performance.

METHOD

Participants

To test these hypotheses, we asked 63 pairs of participants (19 female pairs, 17 male pairs, and 27 mixed pairs; mean age 19, mode 18, maximum 28) to engage in three games across a computer network. Participants were drawn from a population of general psychology students who received class credit for participation. All participants signed informed consent forms and were treated according to APA guidelines.

Design

The experiment manipulated proximity and communication in a fully crossed 2 X 2 design with four true experimental levels: colocated with free talk, colocated with no talk, remote with free talk and remote with no talk. In the colocated treatments, participants sat side-by-side in the same room. In the remote treatments, they sat in separate rooms. In the no-talk treatments they were not allowed to communicate at all. In the colocated free-talk condition, they are allowed to communicate freely. In the remote free-talk condition, they communicated over a radio.

As part of the experiment, pairs of participants in all four conditions made choices that revealed levels of trust. Their choices self-select the pair into one of three categories: trust, distrust, or betrayal. This quasi variable serves as the third factor in our design.

Procedure

Pairs of participants were randomly assigned to one of four conditions defined by two levels of communication (open and none) and two levels of proximity (colocated and remote) and asked to play three computer games. The first was a test of individual cognitive processing, the mental rotation task of Shepard and Metzler (1971). The second task was the first of the two-person team tasks -- the extensive form economic trust game of McCabe, Rassenti, and Smith (1996), shown in Figure 1. The teams then played a distributed command and control task we call the 'Boat Game.' Finally, each participant individually performed a second trial of the mental rotation task.

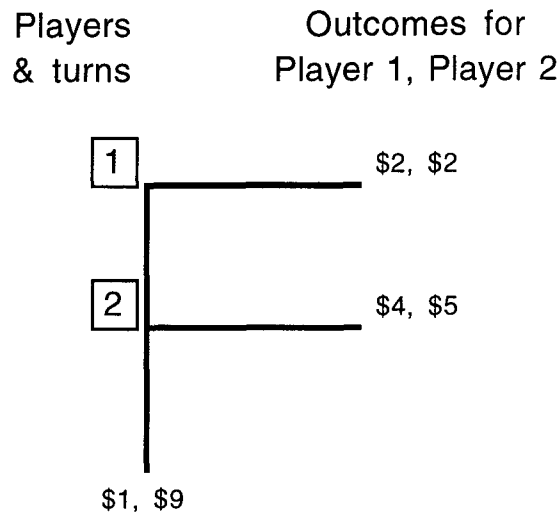


Figure 1 - The extensive-form trust game used as the stimulus in the quasi-manipulation of trust. The numbers in the squares represent Player 1 and Player 2. Player 1 has the first move. A move to the right is a distrusting move that nets both players $2. A move down is a trusting move that passes play to Player 2. A move right by Player 2 is a trustworthy move that nets Player 1 $4 and Player 2 $5. A move down by Player 2 betrays Player 1's trust and awards $9 to Player 2 but only $1 to Player 1.

The mental rotation task exercises the individual's capacity for mental imagery. It presented two three-dimensional objects on a computer screen. Participants were asked to identify whether the objects were the same or different (mirror images). To respond, the participant pressed the S key if the objects appeared to be the same and the D key if the objects appeared to be different. The dependent variables were reaction time and accuracy. This task was administered both before and after the team tasks in a within-subjects (individual) repeated-play design. Learning should allow the participants to improve across repetitions. Any decrement would suggest that the intervening team tasks have a detrimental effect on individual cognitive performance. The software platform for the mental rotation task was SuperLab Pro (© Cedrus Corporation).

The first team game was the trust game shown in Figure 1. In this game, participants in the role of player 1 make a decision to move right or to move down. Moving to the right ends the game, gives both players $2, and places the team in the distrust condition. Moving down passes play to player 2 and is an implicit statement of trust. By moving down, player 1 is trusting player 2 to move to the right, netting both a larger payoff. If player 2 is given the chance to play, he or she may choose to move right or down. Moving right nets player 1 $4 and player 2 $5, ends the game, and places the team in the trust (and trustworthy) condition. Moving down is an explicit betrayal of player 1's trust. This move nets player 2 $9 while player 1 receives only $1. It also ends the game and places the team in the betrayal condition. Each team played the game only once and, in the process, placed itself into one of three categories of trust – trust (and trustworthy), distrust, and betrayal. This self-selection is a participant variable and the resulting treatments are quasi-exp imental variables. The design is accordingly 2 x 2 x 3 (communication by proximity by trust).

The second team game was the Boat Game. The goal of the game is for the team to work together to navigate a tugboat down a river. One participant acted as the pilot, responsible for the steering of the boat, while the other was the engineer, in charge of the speed of the boat. This division of duties makes cooperation an essential ingredient of successful team performance. Like all tugboats, the boat has considerable inertia and takes a while to respond to commands to turn, to speed up, and to slow down. If the boat hit the shore, it would stall for five seconds and then move to the middle of the river at a full stop. The number of 'crashes' is our measure of team error.

There was a limited amount of time and fuel that could be used to reach the end of the river. Two meters that used dollars and cents to represent time and the amount of fuel remaining started at $15.00 each. Every command to increase speed burned fuel and deducted $0.05 from the meter for fuel. Time cost money as well. The meter for time counted down at a rate of $0.01 per second. Elapsed time is our measure of team performance.

The two team members split the total shown on both meters when their boat reached the end of the river. If either meter reached zero before the end of the river, the game ended and both team members received no money. The platform for both the trust game and the boat game was Java script written for this experiment.

RESULTS

By manipulating communication and proximity, and by allowing participants to reveal trusting behavior at three levels, and by measuring performance in the boat game, we test whether communication, proximity, and trust interact to influence team performance in a distributed command and control task. The team is the unit of analysis except for the mental rotation task in which the individual is the unit of analysis. We used a within subject t-test to analyze the mental rotation data and ANOVA to analyze the dependent measures from the boat game.

We had hoped to have a roughly equal number of teams self-select into each of the three categories of trust. Accordingly, the three sets of payoffs in the trust game shown in Figure 1 were pre-tested so that approximately 33% of teams selected each outcome. The distribution obtained in this experiment, shown in Figure 2, was quite different. Very few teams experienced distrust and a disproportionate number trusted and were trustworthy. Because the 2 x 2 x 3 ANOVA was highly sensitive to the small number of distrusting teams, these data were removed from the analysis. The resulting 2 x 2 x 2 ANOVA tests the effect of trust and betrayal (and communication and proximity) but not of distrust.

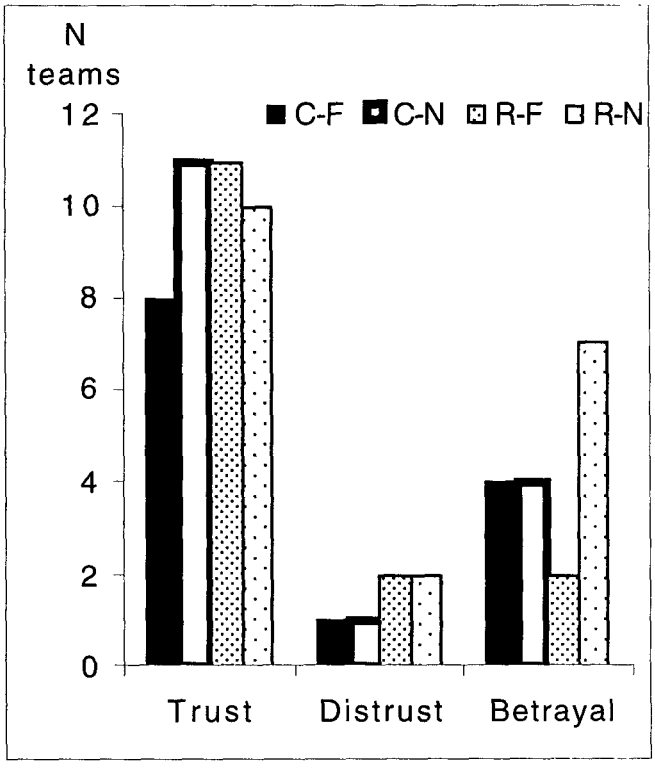


Figure 2. The distribution of teams by trust category. Only six of the 53 teams experienced distrust.

There were no main effects for time on task. The trust-proximity interaction was significant $F(1, 49) = 4.57$, $p < .038$. The data are shown in Figure 3. Betrayal had relatively little effect when team members were in different rooms but seriously hampered coordination when team members were in the same room.

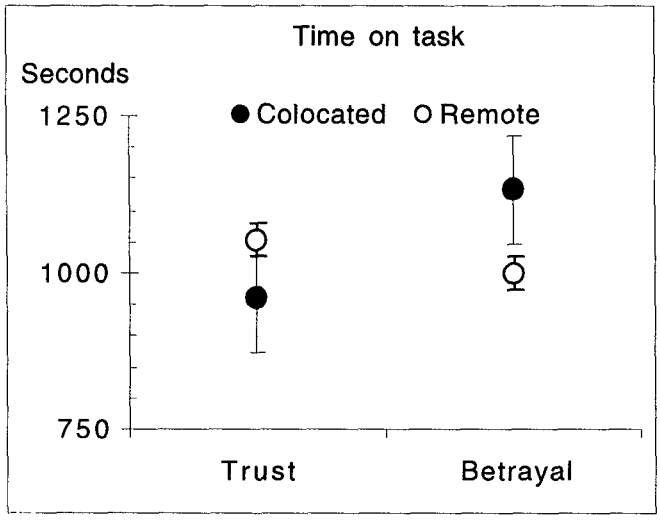


Figure 3. The trust-proximity interaction on time on task.

Figure 3 shows the error data (the number of times the boat hit the bank) aggregated across the two conditions of proximity to show the interaction of trust and communication. The ANOVA reveals that the interaction is pragmatically significant $F(1, 49) = 3.49$, $p < .068$. Teams that experienced betrayal were less effective at controlling the boat if, and only if, they were allowed to communicate. In contrast, betrayal had no discernable effect when teams worked in silence.

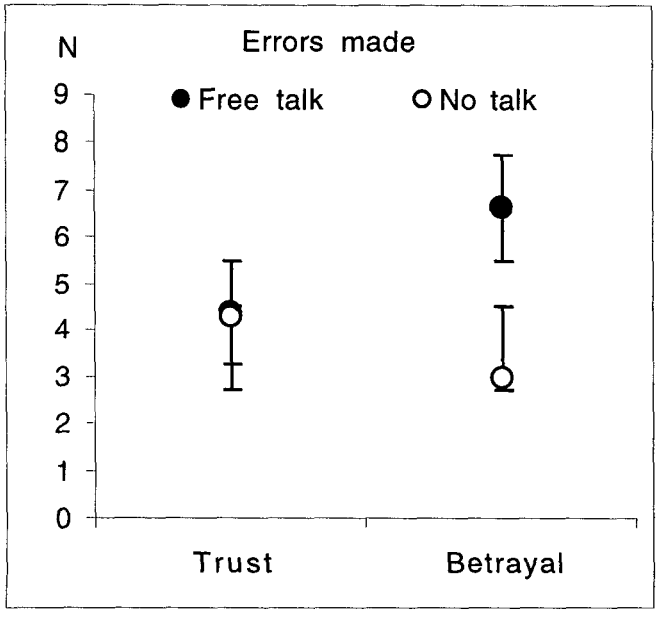


Figure 3. The interaction of trust and communication on errors (the number of times the boat crashed into the bank).

Of the 126 total participants, 112 were both faster and more accurate in the second trial of the mental rotation task. This improvement reflects learning. As expected, the manipulations of communication, trust, and proximity had no effect on this individual cognitive process. This result implies there is a dissociation of the effect of communication, proximity, and trust on team and individual performance.

DISCUSSION

This experiment indicates that the quality of trusting behavior exhibited by team members can influence how they interact and how well the team performs a distributed command and control task. Specifically, betrayal - the failure to reciprocate an expression of trust with trustworthiness - can have a profoundly negative influence on how teams communicate and on how they coordinate their activities when face-to-face.

With transition to the Objective Force Warrior, the Army is committing to distributed command and control. If these findings generalize to a combat environment, the trainers and leaders of these high-tech units will have to find a way to instill trust in and between all team members.

The Army's profound change in organizational structure has analogs in education and industry. Distance learning and video conferencing rely on electronic technology to connect dispersed individuals working together to achieve a common goal. Teachers and managers cannot afford the consequences of misplaced trust. When teams are formed that do not have the solid foundation of mutual trust, they may have trouble communicating and working in close quarters.

ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Dr. Elmar Schmeisser is the Technical Monitor. The Java scripts for the trust game and the boat game were written by Blaine Williams.

REFERENCES

Latané, B. (1981). The psychology of social impact. American Psychologist **36**, 343-356.

Latané, B., & L'Herrou, T. (1996). Spatial clustering in the conformity game: Dynamic social

impact in electronic groups. Journal of Personality and Social Psychology **70** (6), 1218-1230.

Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.) Trust in Organizations. Thousand Oaks, CA: Sage.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. Academy of Management Review **23** (3), 438-458.

McCabe, K. A., Rassenti, S. J., & Smith, V. L. (1996). Game theory and reciprocity in some extensive form experimental games. Proceedings of the National Academy of Sciences, USA **93**, 13421-13428.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Academy of Management Review **23** (3), 393-404.

Shepard, R. N. & Metzler, J. (1971). Science **171** (3972), 701-703.

Yates, J. F., & Stone, E. R. (199 The risk construct. In J. F. Yates (Ed.) Risk Taking Behavior. New York: Wiley.

DISTRIBUTED TEAMS AND REVEALED TRUST

Shala Bloomberg, Serena Hawkins, Rachel Davidson, Benjamin Swan
Kansas State University

Kip Smith
Linköping University/Institute of Technology

This paper discusses a laboratory study of the influence of proximity, communication mode, and performance on the formation of trust between team members. The study manipulated team proximity and communication mode and measured (1) the performance of two-person teams in a distributed command and control task and (2) their subsequent expressions of trust as revealed in an economic game. Measures of team performance included the number of errors made and time on task. Revealed trust was scored at three levels: trust, distrust, and betrayal. We treat performance in the command and control task as a subject variable to determine whether it has an impact on the subsequent expression of trust. We found a weak main effect for communication mode on performance and no effect of performance on revealed trust. However, there was a highly significant main effect for proximity on trust. Members of dispersed teams are less likely to trust each other than members of teams who work in the same location. This finding runs counter to the current trend toward distributed work groups.

INTRODUCTION

Camaraderie has long been the foundation for team cohesiveness and effectiveness. Teams train together, learn together, and grow together. Traditionally, team members maintain close contact, both in terms of physical proximity and direct communication. In the process, they form the bonds of friendship, team identity, and trust. Their effectiveness as a team has been rooted in their open communication, close proximity, and mutual trust. The recent trend towards distributed work groups and teleconferencing flies in the face of this tradition. Camaraderie is downplayed. Teams no longer work face-to-face and often communicate solely via electronic links. It is unknown how this change in organizational structure will effect team members' ability to trust their leaders and each other.

Trust necessarily involves an interaction between two or more individuals or groups (e.g., Costa, Roe, & Taillieu, 2001). For trust to exist, the party that acts first (or whose action is revealed first) must exhibit trusting behavior and the party that responds must prove to be trustworthy. Following Rousseau et al. (1998), we define trusting behavior as the willingness to be vulnerable to the actions of others (the second party, those who are trusted) based upon positive expectations of the others' behavior. As such, trust is more than cooperation. It is cooperation given a palpable likelihood of (opportunity) loss - risk - and an expectation of positive reciprocation (Lewicki, McAllister, & Bies, 1998; Yates & Stone, 1992). Trustworthiness supports behavior that honors the first party's implicit expectation of positive reciprocation. It often, but not necessarily, incurs some measure of opportunity loss. A lack of trust can be manifested in two distinct ways, by either distrust or betrayal. Distrust is marked by the absence of trusting behavior by the party to act first while betrayal is marked by the absence of trustworthy behavior by the party that responds.

Previous research has supported the intuition that team members who work well together tend to trust each other (Lewicki & Bunker, 1996). In contrast, people who have never met or seen each other are more likely to distrust or to betray. We expect that team members who work well together in our distributed command and control task will be more likely to trust each other than the members of teams that do poorly.

Further, we expect communication and proximity to influence the quality of performance and the formation of trust. Teams that are permitted to communicate freely, either face-to-face or over electronic links, should be more efficient and effective than teams that do not communicate. This hypothesis is supported by previous research that has shown that physical proximity and communication mode influence a person's willingness to comply with a request. For example, in the classic study by Milgram (1974), a research participant was far more likely to follow instructions if the communication was direct (e.g., face-to-face) and if the researcher was in the same room. If the researcher communicated from a separate room via telephone, the participant was three times less likely to comply. These findings suggest that

there is good reason to believe that the mode of communication and physical proximity may interact to influence team behavior. This study seeks to isolate the effects of proximity and mode of communication on revealed trust in order to inform the design of training modules for distributed command and control tasks.

In addition to studying the interaction of performance, communication, and proximity and their effect on revealed trust, we are interested in their impact on individual cognitive performance. We have no *a priori* reason to expect these factors to influence individual performance. In particular, we expect to observe no detrimental effect of team proximity, communication mode, performance, or category of trust on individual cognitive performance. Accordingly, we expect to find a dissociation between our measures of team and individual performance.

METHOD

Participants

Participants were 54 pairs of undergraduate students (9 female pairs, 17 male pairs, and 28 mixed pairs) from general psychology classes. Their median age was 19 and the mode 18; the oldest was 40. We did not ask participants about their previous experience working as a member of a distributed team. While it is impossible to ensure that team members did not know each other, every effort was made to keep team membership anonymous prior to the beginning of the experiment. All participants signed informed consent forms and were treated according to APA guidelines.

Design

The experiment manipulated two factors, each at two levels. The first factor was proximity, whether the participants were in the same or different rooms during the experiment. The second factor was mode of communication, whether or not they were allowed to talk during the experiment. In the colocated - free talk condition, the participants were placed in the same room and allowed to converse freely. In the remote - free talk condition, the participants were in different rooms and communicated freely via radio. Remote - no talk participants were located in different rooms and allowed no form of communication. The colocated - no talk participants were placed in the same room but were not allowed to communicate.

As part of the experiment, pairs of participants in all four conditions made choices that revealed levels of trust. Their choices self-selected the pair into one of three categories: trust, distrust, or betrayal. Teams of two are the between-subjects units of analysis.

Procedure

Pairs of participants were randomly assigned to one of four conditions defined by two levels of communication (open and none) and two levels of proximity (colocated and remote). They played three games presented on a network of computers.

The first part of the experiment was a test of individual cognitive processing, the mental rotation task of Shepard and Metzler (1971). It presented two three-dimensional objects on a computer screen. Participants were asked to identify whether the objects were the same or different (mirror images). To respond, the participant pressed the 'S' key if the objects appeared to be the same and the 'D' key if the objects appeared to be different. The dependent variables were reaction time and accuracy. This task was administered both before and after the team tasks in a within-subjects (individual) repeated-play design. Learning should allow the participants to improve across repetitions. Any decrement would suggest that the intervening team tasks have a detrimental effect on individual cognitive performance. The software platform for the mental rotation task was SuperLab Pro (© Cedrus Corporation).

The next part of the experiment was known as the Boat Game. The goal of the game is for the team to work together to navigate a tugboat down a river. One participant acted as the pilot, responsible for steering the boat, while the other was the engineer, in charge of the speed of the boat. This division of duties makes cooperation an essential ingredient of successful team performance. Like all tugboats, the boat has considerable inertia and takes a while to respond to commands to turn, to speed up, and to slow down. If the boat hit the shore, it would stall for five seconds and then move to the middle of the river at a full stop. The number of 'crashes' is our measure of team error.

There was a limited amount of time and fuel that could be used to reach the end of the river. Two meters that used dollars and cents to represent time and the amount of fuel remaining started at $15.00 each. Every command to increase speed burned fuel and deducted $0.05 from the meter for fuel. Time cost money as well. The meter for time counted down at a rate of $0.01 per second. Elapsed time is our measure of team efficiency.

The two team members split the total shown on both meters when their boat reached the end of the river. If either meter reached zero before the boat reached the end of the river, the game ended and both team members received no money.

Following the boat game, the teams played a second game, the extensive-form economic trust game of McCabe, Rassenti, and Smith (1996), shown in Figure 1. The extensive-form game has been used extensively by

experimental economists to assess trust (e.g., McCabe and Smith, 1999; Gunnthorsdöttir, McCabe, and Smith, 2002). In this game, participants in the role of player 1 make a decision to move right or to move down. Moving to the right ends the game, gives both players $2, and places the team in the distrust condition. Moving down passes play to player 2 and is an implicit statement of trust. By moving down, player 1 is trusting player 2 to move to the right, netting both a larger payoff. If Player 2 is given the chance to play, he or she may choose to move right or down. Moving right nets player 1 $4 and player 2 $5, ends the game, and places the team in the trust (and trustworthy) condition. Moving down is an explicit betrayal of player 1's trust. This move nets player 2 $9 while player 1 receives only $1. It also ends the game and places the team in the betrayal condition. Each team played the game only once and, in the process, placed itself into one of three categories of trust – trust (and trustworthy), distrust, and betrayal. This outcome is our main dependent variable.

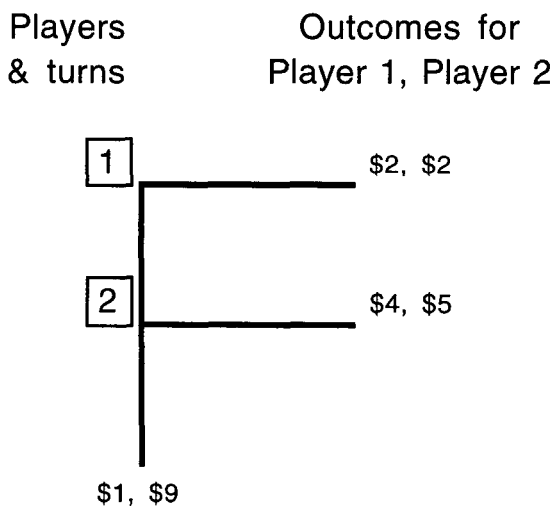


Figure 1 - The extensive-form trust game used as the stimulus in the quasi-manipulation of trust. The numbers in the squares represent Player 1 and Player 2. Player 1 has the first move. A move to the right is a distrusting move that nets both players $2. A move down is a trusting move that passes play to Player 2. A move right by Player 2 is a trustworthy move that nets Player 1 $4 and Player 2 $5. A move down by Player 2 betrays Player 1's trust and nets Player 1 $1 and $9 for Player 2.

In the tradition of experimental economics, participants played for US dollars that were received immediately upon completing the two games. There is extensive literature in economics that supports the use of *real* money rather than virtual money as an incentive (e.g., Camerer and Hogarth, 1999; Gneezy and Rustichini, 2000; Güth and Kliemt, 1994). We did not collect subjective ratings of trust to avoid inducing response bias.

Analysis

The team is the unit of analysis except for the mental rotation task in which the individual is the unit of analysis. We used a within subject t-test to analyze the mental rotation data and a 2 x 2 ANOVA to analyze the effects of proximity and communication on team performance. Because the trust data are categorical dependent measures, the effects of performance, proximity, and communication on revealed trust were analyzed using logistic regression.

RESULTS

This experiment addressed a series of issues related to team performance and trust formation: (1) the impact of mode of communication and team proximity on team performance in a distributed command and control task, (2) the impact of team performance on subsequently revealed trust, (3) the impact of communication and proximity on subsequently revealed trust, and (4) the impact of the team tasks and revealed trust on individual cognitive processing. These issues are discussed in turn.

A 2 X 2 ANOVA revealed that our manipulations of mode of communication and proximity had no effect on the quality of team performance as measured by the number of errors made (crashing the boat into the bank) and time on task. Similarly, the logistic regression using errors and time on task as predictors of trust category suggests that team performance in this command and control task has no effect on revealed trust.

Figure 2 aggregates across the manipulation of communication to show the percentage of teams that self-selected into each category of trust in the two conditions of proximity. The logistic regression fails to reject the null hypothesis of no lack of fit. The Chi-square lack of fit test $\chi^2(2) = 1.31$, $p < .02$ reveals that the main effect of proximity is highly significant. As shown in Figure 2, only one of the teams that worked together experienced distrust. The corresponding test statistic for communication is not significant, $p > .89$, Figure 3. There was no main effect of communication and no interaction between proximity and communication on revealed trust.

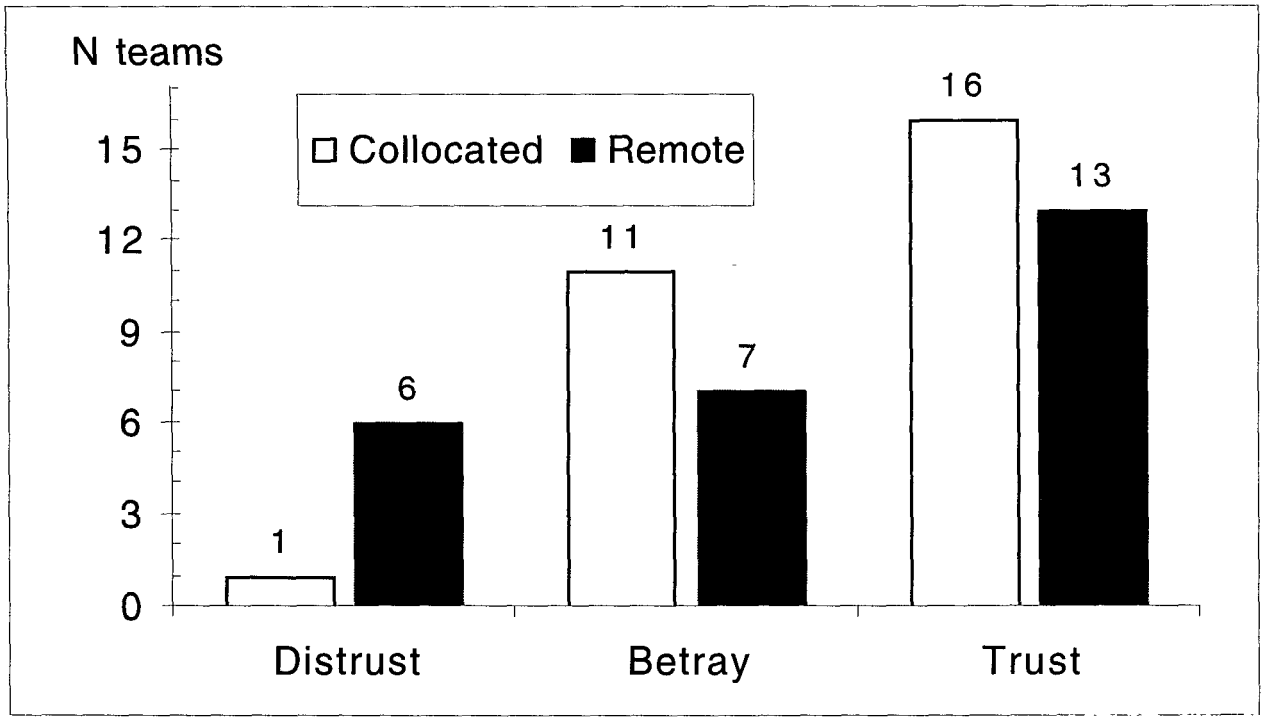


Figure 2 - Graph showing the main effect of the manipulation of proximity on the categorical variable trust. Only one of the colocated teams exhibited distrust.

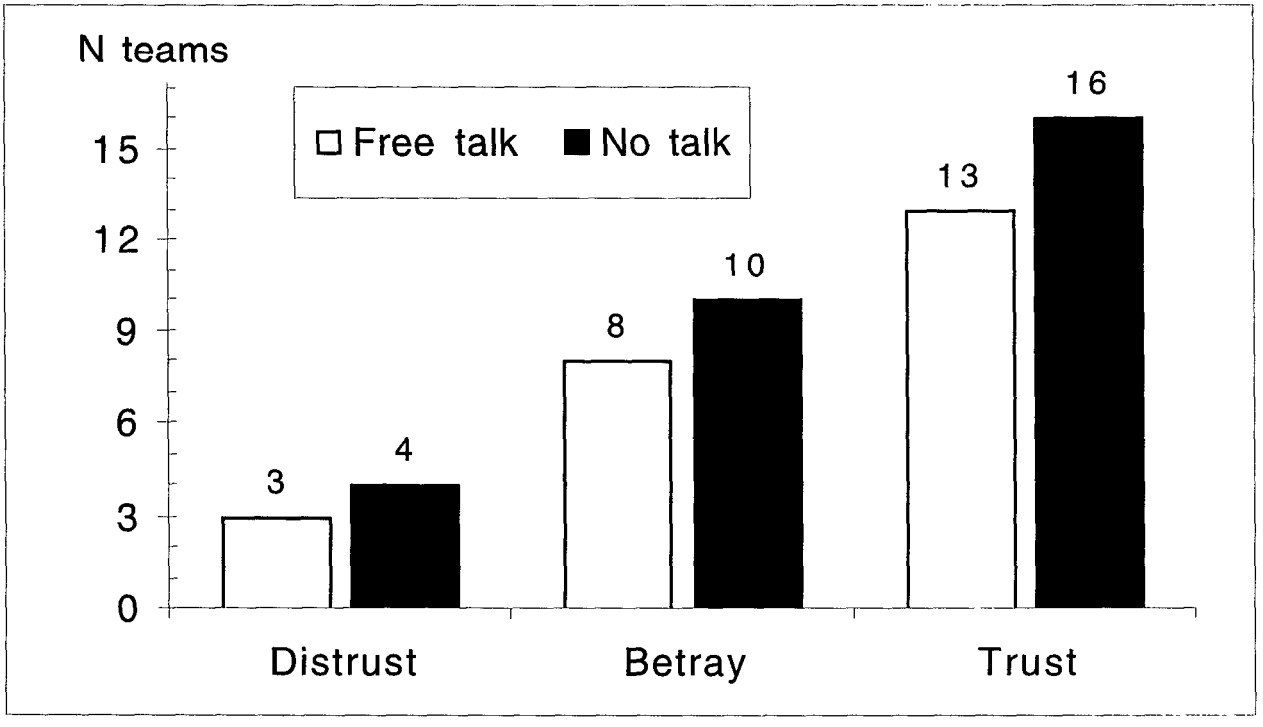


Figure 3 - Graph showing the absence of a main effect of the manipulation of communication on the categorical variable trust.

Of the 108 participants, 97 were both faster and more accurate in the second trial of the mental rotation task. This improvement reflects learning. As expected, the manipulations of communication and proximity and the category of revealed trust had no effect the on this individual cognitive process. This result implies there is a dissociation of the effect of communication, proximity, and trust on team and individual performance.

DISCUSSION

Contrary to our expectations, only proximity played a critical role in the formation of trust between members of our teams. The finding that proximity influences trust between team members does not bode well for organizations that form teams with members in dispersed locations. Our data suggest that when such a team is initially assembled, it would be wise to bring them physically together to mitigate the potential emergence of underlying distrust.

Neither the mode of communication nor team performance influenced revealed trust. The observation that communication mode does not influence team performance and revealed trust suggests that, once a team is established, it can work as effectively using electronic media as face-to-face.

These results are likely to be sensitive to the nature of our distributed command and control task. The boat game fully decomposes responsibilities. The pilot steers; the engineer controls the throttle. There is no hidden complexity; responsibility is clear. To test the generality of our findings, we plan to apply the method used this study to process control tasks in which the division of labor is less structured.

ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Dr. Elmar Schmeisser is the project monitor. The Java scripts for the trust game and the boat game were written by Blaine Williams.

REFERENCES

Camerer, C., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. Journal of Risk and Uncertainty **19** (1-3), 7-42.

Costa, A. C., Roe, R. A., & Taillieu, T. (2001). Trust within teams: The relation with performance effectiveness. European Journal of Work and Organizational Psychology **10** (3), 225-244.

Gneezy, U., & Rustichini, A. (2000). Pay enough or not at all. The Quarterly Journal of Economics, **15** (3), 791-810.

Gunnthorsdöttir, A., McCabe, K., & Smith, V. (1999). Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. Journal of Economic Psychology **23**, 49-66.

Güth, W. & Kliemt, H. (1994). Competition or cooperation: On the evolutionary economics of trust, exploitation and moral attitudes. Metroeconomica **45**, 155-187.

Lewicki, R. J, & Bunker, B. (1996). Developing and maintaining trust in work relationships. In R. Kramer & T. Tyler (Eds.), Trust in organizations. Newbury Park, CA: Sage.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. Academy of Management Review **23** (3), 438-458.

McCabe, K. A., Rassenti, S. J., & Smith, V. L. (1996). Game theory and reciprocity in some extensive form experimental games. Proceedings of the National Academy of Sciences, USA **93**, 13421-13428.

McCabe, K. A., & Smith, V. L. (1999). A comparison of naïve and sophisticated subject behavior with game theoretic predictions. Proceedings of the National Academy of Sciences, USA **97**, 3777-3781.

Milgram, S. (1974). Obedience to Authority. New York: Harper & Row.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Academy of Management Review **23** (3), 393-404.

Shepard, R. N. & Metzler, J. (1971). Science **171** (3972), 701-703.

Yates, J. F., & Stone, E. R. (1992). The risk construct. In J. F. Yates (Ed.) Risk Taking Behavior. New York: Wiley.

BEYOND MILGRAM: AN EXPERIMENTAL STUDY OF LEADER PRESENCE

Keith Pangburn, Steven Freund, Holly Pangburn
Kansas State University

Kip Smith
Linköping University/Institute of Technology

The US Army is changing the organizational structure of the combat strike unit to cover more distance and inflict greater damage with fewer soldiers. These changes involve a shift in the modes of command and control between soldiers and their leaders: leader presence on the battlefield will no longer be necessary. Initial experiences with this concept have raised a concern about the effectiveness and efficiency of remote command and control. In the experiments discussed here, we address this concern in an ecologically realistic simulation of a battlefield environment. We test the impact of leader presence at two levels (present and remote) on combat task performance. Participants' task was to shoot, move, and communicate with a leader. We found that participants were faster to move and shoot in the leader-present condition than in the leader-remote condition.

Introduction

The face of the modern battlefield has undergone radical transformations over the last decade. Instead of being physically present over the battlefield in a conventional attack aircraft, pilots are now remotely operating unmanned aerial vehicles. US Army soldiers who at one time stood side-by-side in battle are now separated by 10 to 50 meters and are equipped with satellite-communication devices that allow immediate contact with generals in Washington. The practice of having soldiers on the battlefield receive orders from afar through electronic means of communication is known as remote command and control. Increasing the distance between individual soldiers on the battlefield and the reliance on remote command and control are two of the cornerstones of the US Army's plan for modernizing the infantry. The soldiers who will operate under this plan are known as Objective Force Warriors. With the advent of the Objective Force Warrior concept, soldiers will no longer take their battle commands from a leader standing within visual range. Instead, the only connection with their commanding officers will be their radios or other electronic information devices.

The change from leaders who are present on the battlefield to leaders who give orders from a distance is likely to have an impact on soldier performance. Designing this organizational change and mitigating the impacts raises human factors concerns. The experiments discussed here investigate the effect of that change on soldiers' response to commands to move and to shoot.

Previous research has shown that varying the physical proximity of an authority figure affects a person's compliance with a command. In the classic study by Milgram (1974), a research participant was far more likely to administer electric shocks to another person at the researcher's command if the researcher was present. If the researcher gave an order to punish an individual from a separate room via telephone, the participant was three times less likely to comply with the command than if the researcher were in the room giving the command.

The organizational change proposed by the US Army is likely to do more than remove the leader from the battlefield. It also has the potential to erode common ground between soldiers and their leader. The essence of common ground is mutual knowledge of a shared representation of information (Clark, 1996; Fairburn & Wright, 2000). Common ground has traditionally been the glue that holds a fighting unit together. To paraphrase:

> When we have common ground, I know what you know and you know what I know. Further, I (you) know that you (I) know what I (you) know. We all know that everyone else knows what we know.

We adopt the concept of common ground to provide the theoretical foundation for this investigation. We believe that common ground works to support a participant's belief in the credibility of a leader's authority. A leader who is in the field with his troops should have more credibility than an absent or remote leader. Accordingly, the use of remote communication can be expected to degrade common ground and the leader's ability to exercise authority. We hypothesize that this degradation in perceived authority will be reflected in slower reaction times when commands are given remotely than when they are given face-to-face. If this is found to be the case, it will be necessary to design interfaces and training

regimes to insure that this degradation of authority can be mitigated.

Recent events in Somalia support the hypothesis that reaction times will be slower when leaders are not present in the field (Bowden, 1999). In this theater of operations where the leadership system was not designed for remote-authority operation, remote leadership and electronic communication contributed to disastrous breakdowns in combat effectiveness.

In the two experiments presented here, we have modified the Milgram task to make it palatable to institutional review boards and to give it sufficient ecological validity to generalize to a military setting. The technology that enables this simultaneous ethical sanitization and realism is called Paintball.

Method

Participants

Forty-two college students, five women and 37 men, drawn primarily from an introductory psychology class, volunteered to spend a Saturday afternoon playing paintball on our simulated assault lane. The average and median age was 19 with a range of 18 to 26. All signed informed consent forms and liability releases. All participants were treated according to APA guidelines.

Design, measures, and task

We manipulated leader presence at two levels (present and remote) in a within-subjects design. We measured the participant's response time to the leader's commands to move and to shoot. We predict slower response times in the leader-remote condition. We expect that the effect of leader presence will not be strong under normal circumstances, but that it will emerge under the rigors of a simulated combat environment.

The simulated combat environment in this study is a paintball assault lane, Figure 1. The lane consisted of eight protective stations behind which the participants could hide. At the end of the lane was a fortified position where a sniper was positioned. The sniper's task was to shoot participants as they moved up the lane. The participants advanced through the lane one at a time. Each participant had two tasks. The first task was to move from station to station up the lane in response to the command to move. The second task was to shoot targets in response to the command to shoot. Statistical analysis used a within-subjects t test of the mean difference in response times between the leader-remote and leader-present conditions. The response times to commands to move and shoot were recorded by an observer using a stopwatch.

In a proof-of-concept study, we ran the experiment at a commercial paintball field in Keats, KS, where two participants ran through the combat assault lane as a "buddy team". We quickly realized that this design confounded common ground with the leader and common ground between the two participants. The data acquired in the proof-of-concept study were marginally significant but will not be discussed further.

In the two experiments presented here, first at Fort Riley and then on the Kansas State University campus, we ran only one participant through the lane at a time. This design allows us to measure the effect of leader presence/absence on participants' reaction time.

For the first experiment, the US Army provided access to the 25 ft x 200 ft building at Range 52, Fort Riley, Kansas, home of the US Army 1st Infantry and an active training center for artillery. We set up our paintball assault lane in this building. As shown in Figure 1, the lane consisted of a staggered series of eight 'stations' behind which the participant could take cover. At the far end of the lane, a 'sniper' hid behind a screen of netting and protective cover. Twenty students participated in the first experiment. Attendance was limited because a one-way trip to Fort Riley took 45 minutes and required passing through a security gate and a variety of active firing range complexes. All told, the experiment took at least four hours of the participants' time.

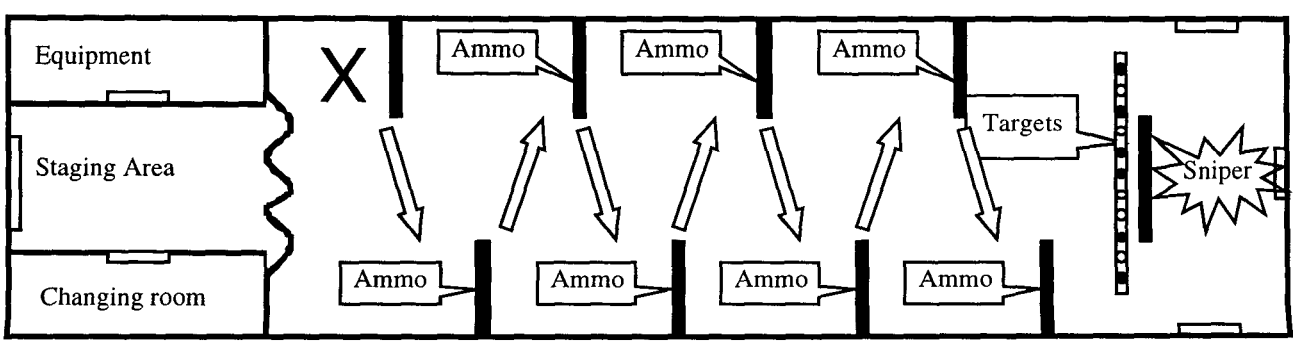


Figure 1 – The layout of the paintball assault lane

In the interest of time and efficiency, we moved the second experiment to an on-campus facility (Quinlan Gardens). The setup was exactly the same as the lane in Figure 1 with one exception. The lane was set up in a small field rather than in a building. Twenty-two students participated in the second experiment.

The change in setting made the leader-remote condition less remote in the outdoor experiment. The first experiment was conducted indoors which allowed the leader in the remote condition to be completely out of sight. The second experiment was conducted outside. In the leader-remote condition, the leader was approximately 100 yards behind the lane but still visible if the participant chose to look. In both experiments, all communication in the remote condition was by radio.

Materials

Participants were given one paintball marker (gun), one set of US Army fatigues, a set of elbow and kneepads, and a paintball face shield. In the leader-remote condition, participants were also given a two-way radio.

Procedure

Upon arriving at our lane a uniformed soldier presented the participants with a standard battle-dress uniform, and asked them to put the uniforms on. The participants were then introduced to their leader, an army officer and the lead author of this paper. The leader briefed them using the official military Operations Order format and addressed them by their last names. They were then told to assemble in a staging area where they could hear the activity in the assault lane while they waited their turns. All of this was purposefully done to immerse the participant in the experiment and to heighten the sense of realism and their anxiety.

While waiting, participants were instructed on the safety and use of the paintball markers and read a briefing. Participants were sent down the lane individually. Whenever the participants took aim at the targets or moved between stations, they exposed themselves to the sniper's fire. Participants were instructed to attempt to shoot enemy targets without hitting friendly targets. No measures were made of firing accuracy, however, because our hypothesis concerns the participant's reaction time to commands given by the leader. The shooting task was created only to give focus to the participant's activity and to give the experiment the feel of a combat environment.

A small container with five paintballs was placed at each of the eight stations. The 40 paintballs in the eight containers were the participant's only ammunition. The participant started at one end of the lane, shown by the X in Figure 1. At this station and all subsequent stations, the leader gave the participant the command "Fire" when he judged it to be safe to do so. The time elapsed from the issue of the command to the first shot fired is the first dependent measure.

When the participants ran out of paintballs, they reported "Out of ammo" to the leader, who then gave the command "Move" when he judged it safe to do so. The participants had to move across the lane to the next station and immediately pick up its container of five paintballs. The time elapsed between the issue of the command to move and the time the participant's hand first touched the new supply of ammunition was the second dependent measure. When the participants finished loading, they reported "Loaded" to the leader who then started the cycle over again with the command "Fire." To measure reaction time, an observer would shadow the participant and record elapsed time with a stopwatch. Although we assume this method was accurate and effective, in the future we hope to record these data using biometric telemetry.

In the leader-present condition, the leader was one station behind the participant. In the leader-remote condition the only contact between the leader and the participant was by two-way radio.

Every participant ran through the lane twice, once with the leader physically present and giving orders, and once with the leader viewing the lane from a remote location and giving orders through a two-way radio. The order of conditions was counterbalanced across participants.

The study was intended to generate some anxiety so that the measures would more readily generalize to the battlefield. The major sources of stress were the fear of being shot and actually being hit by paintballs. The pain associated with being struck by a paintball is slight but real. Protective gear minimizes the risk of injury. A registered nurse received remuneration for being on site throughout the experiment.

Results

The purpose of this study was to test whether we can observe any effect of leader presence on command and control in a simulated combat environment. The participants were exposed to two conditions, leader-present (face-to-face) communication, and leader-remote (two-way radio) communication.

Experiment 1

Figure 2a is a histogram of the within-subject difference in response times to the command to move for the participants in Experiment 1. The difference is calculated by subtracting the average of the times observed in the leader-present condition from the averages of the times observed in the leader-remote condition. Of the 20 participants, 15 took longer to move in the leader-remote condition than in the leader-present condition. A within-subjects t test for these data, $t(19) = 2.958$, $p < .004$, indicates that response times to the command to move were significantly longer in the leader-remote condition.

Figure 2b is the histogram of the within-subjects difference in response times to the command to shoot. The data are, once again, calculated by subtracting the time observed in the leader-present condition from the time observed in the leader-remote condition. Of the 20 participants, 12 took longer to shoot in the leader-remote condition than in the leader-present condition. A within-subjects t test for these data, $t(19) = 2.317$, $p < .016$, indicates that response times to the command to shoot were significantly longer in the leader-remote condition.

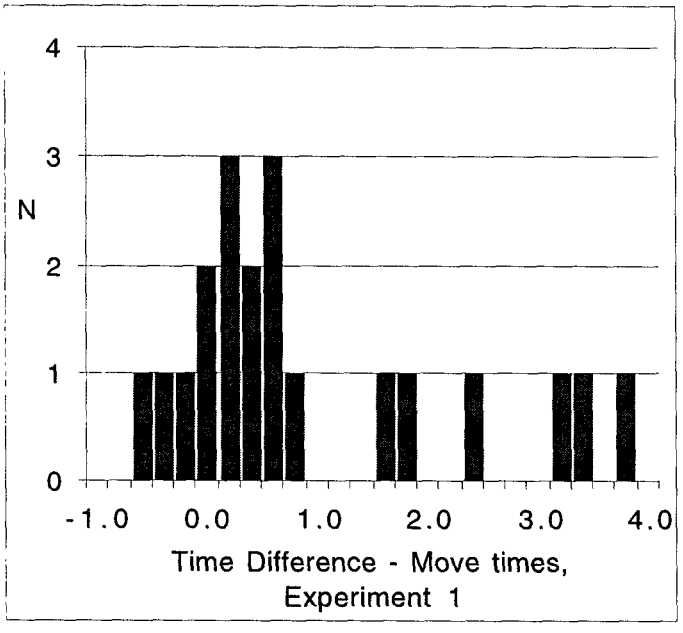
Experiment 2

Figures 3a and 3b are histograms of the within-subjects difference in response times to the commands to move and to shoot, respectively, for the participants in the second experiment. The data are calculated in the same way as those shown in Figure 2. Of the 22 participants, 16 took longer to move in the leader-remote condition than in the leader-present condition. A within-subjects t test for the move data indicates that the difference is significant $t(21) = 2.798$, $p < .006$. Of the 22 participants, 15 took longer to shoot in the leader-remote condition than in the leader-present condition. A within-subjects t test for the fire data indicates that the difference is significant $t(21) = 2.211$, $p < .020$.

Given the similarity of the two experiments' results, it appears the subtle difference in the degree of remoteness of the leader across the two experiments did not have a significant impact on response times. The similarity also allows us to aggregate the data. The test on the composite move data is significant $t(41) = 4.122$, $p < .0001$. The test on the composite shoot data is also significant $t(41) = 3.218$, $p < .0013$. These results support our hypotheses. Participants were faster to react to the leader's commands when the leader was present than when the leader was remote.

Discussion

It is simply not possible to simulate a combat environment in a laboratory. Our desire for ecological validity forced us to work outside the normal laboratory setting. For the first experiment, our assault lane was set in a demolition range building. The lane in the second experiment was set up in a lightly wooded field on campus.



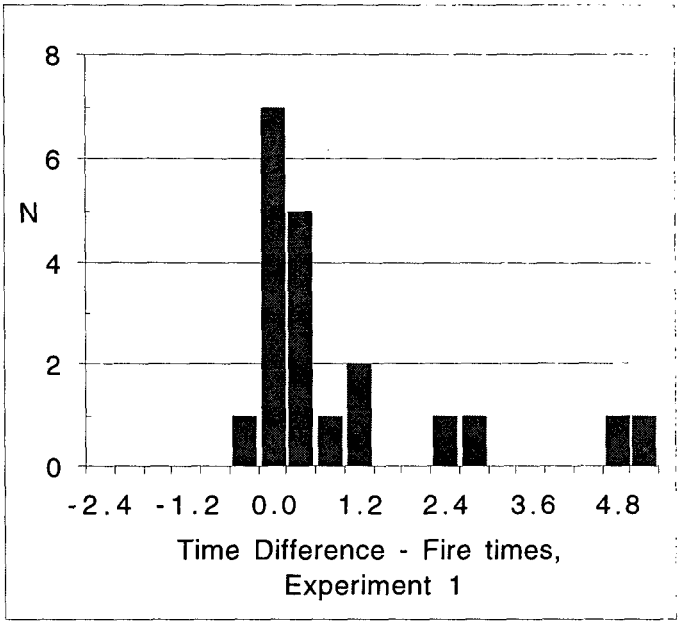


Figure 2 – (A) Histogram of the average differences in participants' response times to the command to move in Experiment 1. Data are calculated as the time to move from one station to the next in the leader-remote condition less the time in the leader-present condition. (B) Histogram of the average differences in participants' response times to the command to fire In Experiment 1.

Embedding experimental controls in a simulated combat environment is a new concept in human factors research. To attain ecological validity, we immersed participants into a pseudo-military environment by providing military uniforms, speaking in military command style, and by keeping the age group in line with that of the military. This procedure gives us the ability to generalize the results of the experiments to real-world battlefield settings. While a paintball lane is not actual combat, we believe our setup comes as close as possible without exposing participants to serious danger. Because the setup and procedure are ecologically valid, we did not need a large sample size.

The two experiments reported here are the first in a planned series of studies on leader presence and common ground. Data acquisition will continue using automated biometric telemetry to eliminate the possibility of observer error in the measurement of response times and to measure psychophysiological stress. Even though organizing the experiment is complex, the procedures and measures are simple and direct. We have shown that a paintball assault lane is an excellent natural laboratory for investigating the impact of leader presence on performance and efficiency within a combat environment.

Acknowledgements

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DOD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Dr. Elmar Schmeisser is the project monitor. We thank the many people who have helped make this experiment possible. A partial list includes Dr. Laurel Allender, Ms. Diane Ungvarsky, Lt. Col. Arthur DeGroat, Dr. Lee Bridenstein, Mr. Andy Kearney, Ms. Kim Jackson, Ms. Jessica McGinnis, Dr. E. J. Reppert, Dr. Rick Scheidt, and the owner and staff of Krazy Kris' Paintball Range, Keats, Kansas.

References

Bowden, M. (1999). Black Hawk Down. New York: Signet Press.

Clark, H. H. (1996). Using Language. Cambridge, UK: Cambridge University Press.

Fairburn, C. & Wright, P. (2000). Exploring the metaphor of automation as a team player. Proceedings of the 10th European Conference on Cognitive Ergonomics. Linköping, Sweden.

Milgram, S. (1974). Obedience to Authority. New York: Harper & Row.

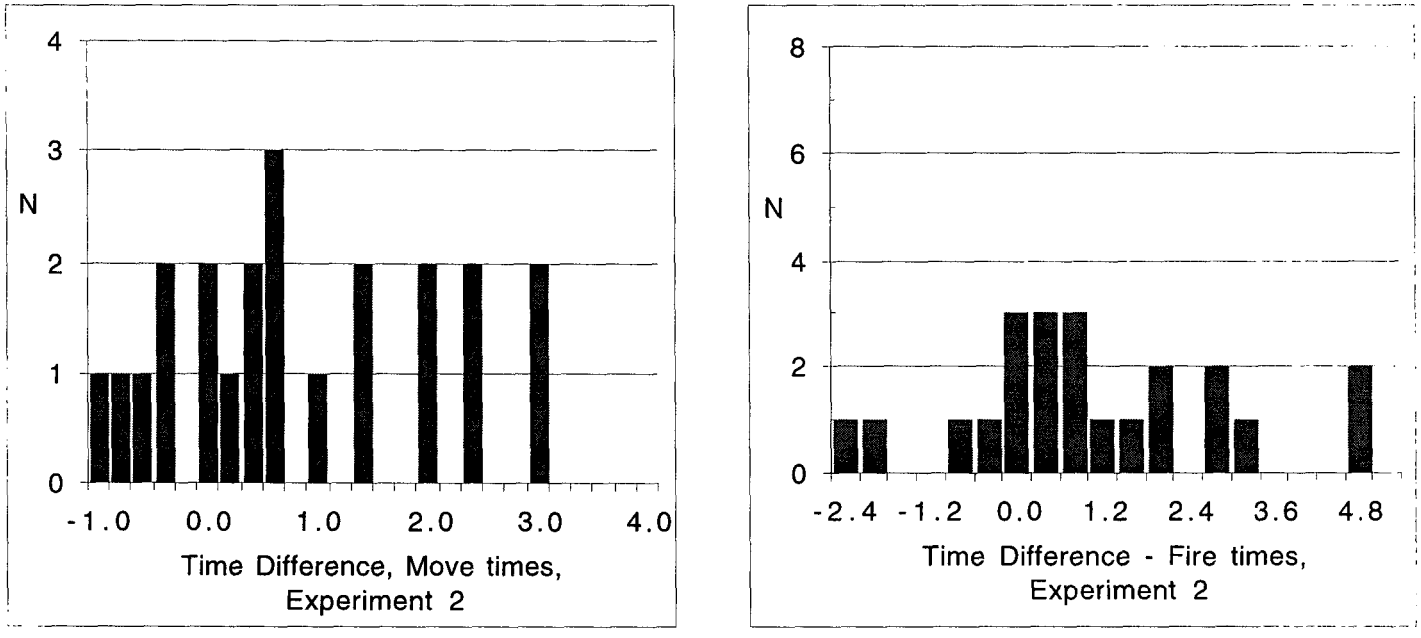


Figure 3 – (A) Histogram of the average differences in participants' response times to the command to move in Experiment 2. (B) Histogram of the average differences in participants' response times to the command to fire In Experiment 2.

Human Performance, Situation Awareness and Automation Technology

Topic Areas

- Team Performance
- Military Systems
- Stress

Submission of 500 word Abstract

Initial experiments on cultural factors that may mitigate the adverse effect of distrust on team cognitive performance under live fire

We are conducting a series of field and laboratory experiments designed to uncover cultural factors that may mitigate the adverse effects of distrust on individual and team performance under live fire. This paper reports the results from four experiments conducted in the United States and discusses experiments that are in progress in Sweden.

The two field experiments conducted in the United States revealed the adverse impact of remote command and control on individual performance under live fire. The venues for these experiments were paintball assault lanes at and near Fort Riley, Kansas. In both experiments, leader presence was a within-subject variable at two levels, leader-present and leader-remote. In the leader-present condition, the leader moved down the assault lane with the participant. In the leader-remote condition the leader monitored the participant's progress from a remote location and communicated using a radio. In both conditions, the participants' task was to obey the leader's commands to move down the lane and to shoot while under live paintball fire from a sniper. Participants' response times to commands to move and to shoot were significantly longer in the leader-remote condition, $t(41) = 4.122$, $p < .0001$ and $t(41) = 3.218$, $p < .0013$, respectively. Participants were faster to obey the leader's commands when the leader was present than when the leader was remote.

The two U.S. laboratory experiments documented the adverse effect of distrust on team performance in a distributed command and control task and a distributed problem solving task. In the laboratory experiments, teams of two played an economic trust game to self-select into conditions of trust, distrust, and betrayal prior to performing a distributed command and control task. Teams that were trusting (and trustworthy) successfully completed the problem solving task in less time than teams that experienced distrust, $F(2, 53) = 4.52$, $p < .016$.

Experiments in progress in Sweden are replicating those completed in the United States and will extend them by combining the experimental venue of the paintball assault lane, mission-based fire-team training scenarios developed by the Army Research Institute, and the economic quasi-manipulation of trust. We plan to assess observed contrasts between Swedish and American teams using the work of Sutton and her colleagues at the Army Research Laboratory, Fort Sill, Oklahoma. The template suggests three dimensions along which culture might influence team cognitive performance: tolerance for hierarchical authority, risk aversion, and social coherence.

Compared to Americans, Swedes tend to prefer flatter hierarchies of command and control, and to be more risk averse and more interdependent. These traits suggest that, given identical command and control problems, ad hoc teams of Swedes are likely to be more forgiving of remote leaders but less effective when experiencing distrust than their American counterparts.

Author Response Form

Title of Submission:
Initial experiments on cultural factors that may mitigate the adverse effect of distrust on team cognitive performance under live fire

This submission is a : Paper __XX__ Poster ____ Panel Discussion Topic ____

Author:
Is this author the point of contact for this submission? Yes __XX__ No ____
Title: Dr. __XX__ Mr. ____ Ms. ____

Name:	Kip Smith
Position / Title:	Research professor
Company:	Linköping Institute of Technology
Department:	Industrial Ergonomics
Address 1:	SE 581 83 Linköping
Address 2:	Sweden
Email:	kipsm@ikp.liu.se

Please check as appropriate:
1. I will be attending this conference __XX__
2. If this is a paper or poster, I will be presenting this paper or poster __XX__
3. I would like to participate as a Chair or Co-Chair for a paper, panel, or poster session if the opportunity is available ____
4. I would like to participate as a paper/panel/poster reviewer if the opportunity is available __XX__

Revealed trust and team performance

Kip Smith

Division of Industrial Ergonomics

Linköping Institute of Technology

581 83 Linköping

Sweden

The US Army is changing the organizational structure of the combat strike unit to cover more distance and inflict greater damage with fewer soldiers. These changes involve a shift in the modes of command and control (a) between soldiers and their leaders and (b) within combat units. Soldiers on the battlefield will receive orders from afar through electronic means of communication. This practice is known as 'remote command and control.' Reliance on remote command and control has the potential to erode the common ground and trust that have long been the foundation of effective fighting units. At issue are (1) whether trust and distrust (and betrayal) have differential effects on team performance and (2) the design of interventions that seed the formation of trust.

We have addressed the first issue in a pair of quasi-experiments that use the extensive-form trust game (e.g., McCabe, Rassenti, and Smith, 1996) as an independent variable. The trust game has three possible outcomes: distrust, trust/trustworthiness, trust/betrayal. In our laboratory, teams of two played the trust game and immediately received cash payoffs. By playing the game, the teams self-selected into one of three conditions in our quasi-experiment: distrust, trust, or betrayal. The teams then engaged a second task that required teamwork and cooperation. Dependent variables were measures of team performance in the second task.

In the first quasi-experiment, the second task was a team problem-solving task. Teams that were trusting (and trustworthy) successfully solved the problem in less time, $F(2, 53) = 4.52$, $p < .016$, than teams that experienced distrust and betrayal. In the second experiment, the second task involved distributed commandand control. Trust interacted significantly with communication mode (face-to-face vs. via radio) to influence the number of errors the team made. The adverse effect of betrayal was felt more strongly by teams that communicated than by teams that did not communicate.

If these results generalize to military tasks, it will be imperative to design interventions that seed the formation of trust. The extensive form trust game appears to be an effective experimental tool for investigating the differential impact of revealed trust on subsequent team performance. It offers the distinct advantage of eliminating the need for experimental manipulation of factors that influence the formation and expression of trust.

Kansas State University
Final Progress Report covering the period 1 January 2003 – 30 June 2003
MURI OPUS -- DAAD19-01-1-0621

Principal Investigator: Kip Smith, Ph.D.
Department of Psychology

A time of transition

This is the final progress report to be filed from Kansas State University (K State) as part of the MURI OPUS research program. The Principal Investigator, Dr. Kip Smith, is leaving K State and joining the faculty of the Division of Industrial Ergonomics at Linköping University/Institute of Technology in Linköping, Sweden. K State has written the University of Central Florida (UCF) that it no interest in pursuing the MURI OPUS work. Linköping and Dr. Smith have submitted a research plan and budget for the second half of 2003 and calendar year 2004 to UCF. That plan builds upon the K State MURI OPUS research that is documented here.

The issue

The US Army is changing the organizational structure of the combat strike unit to cover more distance and inflict greater damage with fewer soldiers. These changes involve a shift in the modes of command and control (a) between soldiers and their leaders and (b) within combat units. Soldiers on the battlefield will receive orders from afar through electronic means of communication. This practice is known as remote command and control.

Reliance on remote command and control has the potential to erode camaraderie and common ground (a) between soldiers and their leaders and (b) within combat units. The essence of common ground is mutual knowledge of a shared representation of information (Clark, 1996; Fairburn & Wright, 2000). Common ground (1) reduces the uncertainty of others' actions, motives, and plans, (2) supports the development of positive expectations about others' actions, motives, and plans, and (3) is one of the cornerstones for the development of trust. Trust is the glue that holds a fighting unit together. The proposed change in organizational structure has the potential to undermine common ground and, hence, trust and, as a result, have a negative impact on the unit's cognitive performance.

The goals of the MURI OPUS research program conducted at K State were to develop methods and models to predict and, ultimately, to mitigate any adverse effects of the loss of common ground and the erosion of trust on individual and team cognitive performance.

Trust, common ground, and cognitive performance

Trust is not a behavior, rather it is a psychological state that influences the interaction between two or more individuals or groups (Berg, Dickhaut, & McCabe, 1995). Following Rousseau et al. (1998), we define trust as the willingness to be

vulnerable to the actions of others based upon positive expectations of the others' behavior. As such, trust is more than cooperation; it is cooperation given (1) a palpable likelihood of risk (Yates & Stone, 1992), and (2) an expectation of positive reciprocation (Lewicki, McAllister, & Bies, 1998). The basis for developing an expectation of positive reciprocation is the shared representation of information known as common ground.

Two conditions are necessary for trust to exist, risk and interdependence (Lewicki & Bunker, 1996; Rousseau, et al., 1998). In the absence of risk, there is no need to trust. Interdependence is the need to rely on another to achieve a desired risky outcome. Interdependence creates the environment that gives trust the opportunity to emerge.

Trust and trustworthiness are not the same. Trusting and trustworthy behavior occur in sequence. The individual or unit that is trusting acts first by making itself vulnerable to others (e.g., to its leaders or other team members). It is willing to trust because it expects those others to share its knowledge, motives, and plans. In contrast, the individual or unit that is trustworthy responds by honoring the expectation that is implicit in trust. Trustworthiness often, but not necessarily, incurs some measure of opportunity loss.

A lack of trust can be manifested in two distinct ways, either as distrust or as betrayal. Distrust is marked by the absence of trusting behavior by the party to act first while betrayal is marked by the absence of trustworthy behavior by the party that responds (Burnham, McCabe, & Smith, 2000).

Accomplishments

The trust game

In 2002, we created a cognitive task, the extensive form game "trust game" discussed in Appendix A, that allows two-person teams to reveal their dispositions to trust, distrust, be trustworthy, or betray.

Three laboratory quasi-experiments

In the second half of 2002 and the first half of 2003, we conducted a series of three experiments using the trust game to document the task-specific adverse impacts of distrust and betrayal on team cognitive performance. The first experiment, called Hobbits and Orcs, documented the impacts on team problem solving (Pruitt, Bolton, Kearney, Swan, & Smith, 2003). In this experiment, participants first played the trust game and then worked together to solve a problem. Teams that were trusting (and trustworthy) successfully solved the problem in less time, $F(2, 53) = 4.52$, $p < .016$, than teams that experienced distrust and betrayal.

The second experiment used the same procedure to document the impacts of distrust and betrayal on a distributed command and control task called the Boat Game (Swan,

Younger, Smith, Toll, & Smith, 2003). In his experiment, trust interacted significantly with communication mode (face-to-face vs. via radio), $F(1, 49) = 3.49$, $p < .068$, to influence the number of errors the team made. The adverse effect of betrayal was felt more strongly by teams that communicated than by teams that did not communicate. It appears that many of our participants had trouble talking with someone who recently betrayed them (or whom they had recently betrayed).

The first two experiments were technically quasi-experiments because the teams' play in the trust game was a subject variable and the basis for assigning the team to one of three trust conditions: trust, distrust, or betrayal.

The third experiment flipped the order of the Boat Game and the trust game to examine the effect of cognitive performance on subsequently revealed trust. Team performance in this command and control task had no effect whatsoever on revealed trust (Bloomberg, Hawkins, Davidson, Swan, Smith, 2003).

This suite of experiments established the directionality of the causal relationship between revealed trust and team cognitive performance. Distrust and betrayal were shown to have task-specific adverse effects on team cognitive performance. In contrast, performance appears to have no effect on revealed trust. These results support our hypothesis. We have published three articles on these experiments (Bloomberg, et al., 2003; Pruitt, et al., 2003; Swan, et al., 2003). They are attached as PDF files.

Two field experiments

During the first half of 2003, we conducted two field experiments to test whether we can observe an effect of leader presence on command and control in a simulated combat environment (Pangburn, Freund, Pangburn, & Smith, 2003). The simulated combat environment was a paintball assault lane, shown in Figure 1. Each participant had two tasks: (1) to move from station to station up the lane in response to the leader's command to move and (2) to shoot targets in response to the leader's command to shoot. At the end of the lane was a fortified position where a sniper was positioned. The sniper's task was to shoot participants as they moved up the lane. Participants were exposed to two conditions of communication mode, leader-present (face-to-face) and leader-remote (two-way radio). This procedure allowed us to contrast participants' response times across a control condition and a remote command-and-control condition. A within-subjects t-test found that participants were significantly faster to follow to the leader's commands to move, $t(41) = 4.122$, $p < .0001$, and to shoot, $t(41) = 3.218$, $p < .0013$, when the leader was present than when the leader was remote. While a paintball lane is not actual combat, we believe our setup comes as close as possible without exposing participants to serious danger. We have published an article on this series of experiments (Pangburn, et al., 2003). It is attached as a PDF file.

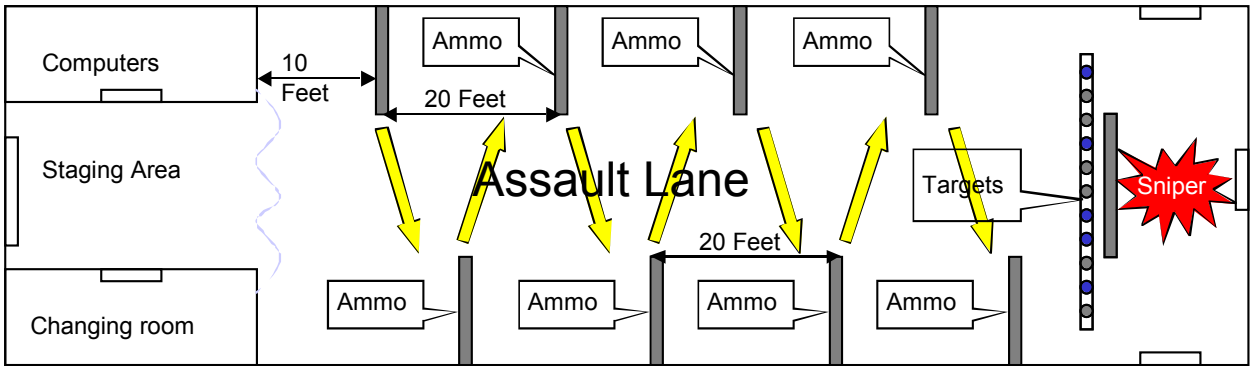


Figure 1. The paintball assault lane used in the K State field experiments.

Implications for the Army

The transition to remote command and control and the advent of the Army of One have the potential to erode the camaraderie and common ground that have traditionally been the foundation of a fighting unit's mutual trust and effectiveness. In addition, the proposed change in organizational structure may remove the leader from the battlefield. These changes raise a pair of questions which the K State MURI OPUS experiments were designed to address:

1) What is the effect of trust, distrust and betrayal on a fighting unit's cognitive performance?

The trust game provides controlled experimental conditions in which teams can either trust, distrust, or betray. This makes it possible for us to create conditions like those that are likely to surround the formation of fighting units composed of the Army of One: conditions in which trust may or may not materialize. Our results indicate that teams that experience trust tend to perform better than teams that experience distrust or betrayal and that the negative effects of distrust and betrayal vary across cognitive tasks.

2) What is the effect of a remote leader on a fighting unit's cognitive performance?

The simulated combat conditions of the paintball assault lane makes it possible for us to ascertain whether or not leader presence is likely to influence the cognitive performance and effectiveness of fighting units. It most definitely does. Participants respond more quickly to commands to move and to shoot when the commands are given by a leader in the field.

Summary and a look ahead

The K State MURI OPUS research has documented a pair of findings that directly relate to the proposed transition to remote command and control and the Army of One. We have identified two sources of stress that are likely to be encountered on the electronic

battlefield and have documented their adverse effects on team cognitive performance. First, a lack of trust appears to be a critical source of stress for teams engaged in cognitive work. Second, a remote leader is obeyed more slowly than an present leader.

The next step in this stream of research is to find interventions that the Army can use to mitigate these adverse effects. The research plan submitted by Linköping University/Institute of Technology (1) builds upon the K State findings and methods, (2) develops hypotheses about an intervention that may mitigate the adverse effects of a lack of trust and remote leaders, and (3) outlines a series of experiments to test those hypotheses.

References

Berg, J. E., Dickhaut, J. W., & McCabe, K.. (1995). Trust, reciprocity, and social history. Games and Economic Behavior **10**, 122-142.

Bloomberg, S., Hawkins, S., Davidson, R., Swan, B., & Smith, K. (2003). Distributed teams and revealed trust. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society. Denver, CO. Oct 13 - 17.

Burnham, T., McCabe, K., & Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. Journal of Economic Behavior and Organization **43**, 57-73.

Clark, H. H. (1996). Using Language. Cambridge, UK: Cambridge University Press.

Cohen, M., Jaffray, J., & Said, T. (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and losses. Organizational Behavior and Human Decision Processes **39**, 1-22.

Fairburn, C. & Wright, P. (2000). Exploring the metaphor of automation as a team player. Proceedings of the 10th European Conference on Cognitive Ergonomics. Linköping, Sweden.

Lewicki, R, J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.)., Trust in organizations: Frontiers of theory and research. Thousand Oaks, CA: Sage.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. Academy of Management Review **23** (3), 438-458.

Pangburn, K., Freund, S., Pangburn, H., & Smith, K. (2003). Beyond Milgram: An experimental study of leader presence. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society. Denver, CO. Oct 13 - 17.

Pruitt, N., Bolton, J., Kearney, A., Swan, B., & Smith, K. (2003). Hobbits and Orcs: communication, trust, and team problem solving. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society. Denver, CO. Oct 13 - 17.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. The Academy of Management Review **23** (3), 393-404.

Smith, V. L. (1982). Research in Experimental Economics **2**. Greenwich CT: JAI Press.

Smith, V. L. (1991). Research in Experimental Economics **4**. Greenwich CT: JAI Press.

Swan, B., Younger, E., Smith, J., Toll, M., & Smith, K. (2003). Communication, proximity, and trust in distributed command and control. Proceedings of the 47th

Annual Meeting of the Human Factors and Ergonomics Society. Denver, CO. Oct 13 - 17.

Tversky, A. & Fox, (1995). Weighing risk and uncertainty. Psychological Review **102** (2), 269-283.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty **5**, 297-323.

Yates, J. F., & Stone, E. R. (1992). The risk construct. In J. F. Yates (Ed.) Risk-Taking Behavior. New York: Wiley.

Appendix A – The extensive form trust game

Cognitive and experimental psychologists have adopted and modified expected utility theory to study the rationality of the decision behavior of individuals (e.g., Curley & Yates and Stone, 1992; Tversky & Kahneman, 1992). These studies typically use paper and pencil tasks to pose hypothetical situations to individuals in a laboratory setting (e.g., Cohen, Jaffray, & Said, 1987; Tversky & Fox, 1995). In contrast, experimental economists have been studying the rationality of interactive decision behavior (e.g., V. Smith, 1982, 1991). These studies typically ask pairs of individuals to play computerized "games" for cash. The games (1) pose opportunities to take or refuse certain gambles or risks involving the use of real money and (2) are usually structured so that the rationality of a decision made by one participant is contingent upon the decision made by the other (e.g., Berg, Dickhaut, & McCabe, 1995).

"Extensive form games" are two-person games that are structured so that both players must think about the other's choices while making their own choices. Play is sequential. The sequential play in the extensive form game enables players to send clear messages: either "I am trusting you. Please be trustworthy." or "I do not trust you." or "I am trustworthy" or "I am betraying your trust."

The figure below presents a simple extensive form game. Play starts after both players have full knowledge of the structure, rules, and possible outcomes (payoffs) of the game. Player A has the first move. She may move right or down. Moving right earns her $1.00 and earns Player B $1.00. Moving down passes play to Player B. Player B may move right or down. Moving right earns Player B $2.50 and Player A $1.50. Moving down earns Player B $4.00 and Player A nothing.

Expected utility theory predicts the moves both players should make. Player B, should she get the opportunity to play, has the choice between payoffs of $2.50 or $4.00. The rational choice is to move down. Knowing that and applying backward induction, Player A should choose to move right. The game is often played repeatedly, with players switching roles between games. Some people in the role of Player A choose to move down, trusting Player B to seize the chance to reward her by moving right. Most people in the role of Player A move right, revealing a distinct lack of trust in Player B.

The sequential play and contingency of behavior in extensive form games creates an environment that puts the disposition to trust on trial. Sequential play makes Player A's commitment to trust explicit to Player B and makes Player B's trustworthiness explicit to Player A. The players have a forced choice: they can be trusting (trustworthy) or not. The behavior cannot be faked, coached, or misrepresented. This relative resistance to faking and the explicit expressions of trust and trustworthiness leads us to hypothesize that the extensive form game is an ideal experimental platform for the evaluation of differences in individuals' dispositions to trust.

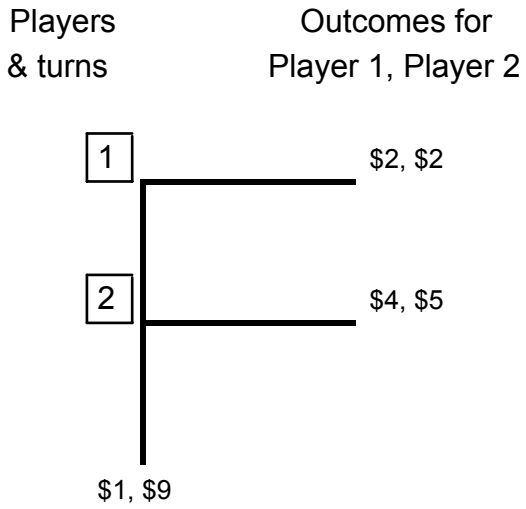


Figure A - The extensive-form trust game used as the stimulus in the quasi-manipulation of trust in the experiments conducted at Kansas State University as part of the MURI OPUS research program. The numbers in the squares represent Player 1 and Player 2. Player 1 has the first move. A move to the right is a distrusting move that nets both players $2. A move down is a trusting move that passes play to Player 2. A move right by Player 2 is a trustworthy move that nets Player 1 $4 and Player 2 $5. A move down by Player 2 betrays Player 1's trust and awards $9 to Player 2 but only $1 to Player 1.

MURI OPUS Research at Kansas State University (K State)

First Half 2003 Progress Report & Final Report

Revealed trust as a source of stress on the electronic battlefield

June 23, 2003

Kip Smith, Ph.D.
Principal Investigator

Contents

1 Final K State two-column progress report:
All goals for first half 2003 have been met.

2 Proposal for second half 2003:

A) Move the K State subaward and research to Linköping University/Institute of Technology, Sweden (LiU).

B) Continue and extend the K State research program at LiU.

3 Summaries of the K State MURI OPUS research program.

A) The effect of leader presence and communication mode on soldier performance under paintball fire.

B) Three laboratory studies on the interaction of revealed trust, proximity, and communication mode on team cognitive performance.

© Kip Smith, 2003

1) Final report on K State MURI research

Goals for Jan - June 2003 from the Dec 2002 K State progress report:

- Paintball: Conduct experiment at Fort Riley (pending ARO approval).
- 3 lab experiments: complete and submit articles for publication.
- Submit abstracts for presentations at HFES & APS.
- Complete new laboratory experiment on the development of trust and teamwork in a virtual world.
- Develop additional field and laboratory experiments on trust and team performance.

All goals met:
Status as of June 25, 2003:

- ✓ Paintball: Two experiments completed. Article attached and summarized here.
- ✓ All three laboratory experiments completed. Three articles attached and summarized here.
- ✓ Four articles written and accepted for presentation at HFES. Poster presented at the American Psychological Society.
- ✓ Virtual world experiment ready for experimental trials at LiU.
- ✓ Modifications of Army STX lanes to fit paintball paradigm ready for deployment at commercial facilities in Linköping.

© Kip Smith, 2003

2) Summary of proposal for second half of 2003

A) Move the K State subaward to Linköping University / Institute of Technology (LiU), Sweden.

- Kip Smith Ph.D., the Principal Investigator for the K State MURI OPUS program, has accepted a research professorship in the Division of Industrial Ergonomics at the Linköping Institute of Technology (LiU) in Sweden.
- K State does not wish to retain the subaward or to continue the MURI OPUS research program.
- LiU and Prof. Smith have submitted research and cost proposals to UCF to move the subaward and research program to LiU.

B) Continue and extend the research program at LiU. Goals for the second half of 2003:

- Recreate and expand the paintball and laboratory experiments at LiU.
- Add psychophysiologic data collection for assessments of stress.
- Refine experiments designed to test hypotheses about the centipede game as a mitigator of stress.
- Establish research links with the Swedish military and other institutions.
- Extend the paintball paradigm from trust-by-individuals-in-a-leader to trust-by-multi-person-teams-in-a-leader and to trust-among-team-members.

© Kip Smith, 2003

3) Summaries of the K State MURI OPUS research program

A) The effect of leader presence and communication mode on soldier performance under paintball fire.

B) Three laboratory studies on the interaction of revealed trust, proximity, and communication mode on team cognitive performance.

© Kip Smith, 2003

3A) The effect of leader presence and communication mode on soldier performance under paintball fire

The Paintball Experiments

Experiment 1: Range 52, Fort Riley
Experiment 2: Quinlan Gardens, K State campus

© Kip Smith, 2003

The paintball experiments 1

Question:

Does leader presence on the battlefield influence the efficiency and effectiveness of a combat unit?

Hypothesis:

The physical proximity of a commanding officer speeds the soldier's compliance with a command.

Task:

Shoot, move, and communicate with the leader while under fire from a sniper shooting paintballs. Move up the assault lane to capture flags at the sniper's hardpoint.

Dependent variables:

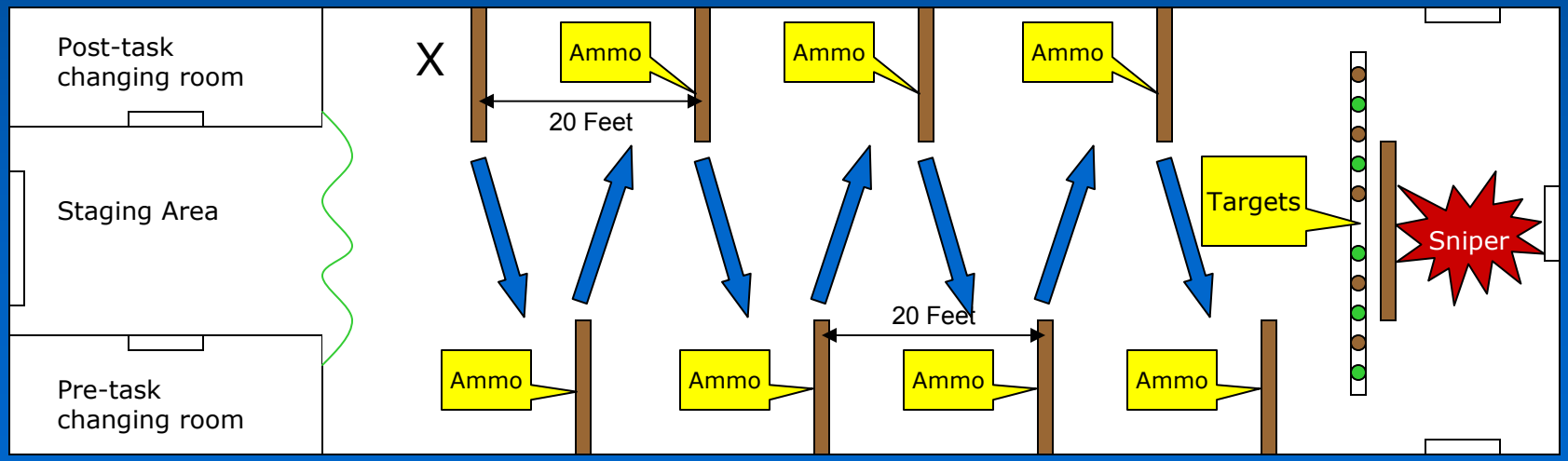
Participant response times to the leader's commands to shoot and to move up the lane.

© Kip Smith, 2003

The paintball experiments 2

Experimental Design:

- 2 factor within-subjects design.
  - Control condition: leader-present - on the lane communicating face-to-face.
  - Experimental condition: leader-remote - off the lane communicating by radio.
- Participants move up the lane twice, once in each condition. Presentation order was counterbalanced across participants.



© Kip Smith, 2003

The paintball experiments 3

The design and layout of the two experiments were identical with the following exception: Experiment 1 was conducted indoors. Experiment 2 was conducted outdoors.

Experiment 1: Range 52, Fort Riley, Kansas.

Indoor lane. In the leader-remote condition, the leader was outdoors, completely out of the participant's sight.

Participants: 20 students, 2 female, age 18 - 27.

Experiment 2: Quinlan Gardens, K State campus.

Outdoor lane. In the leader-remote condition, the leader was ± 50 m behind the lane, out of the participant's line of sight but visible if the participant were to turn around and look.

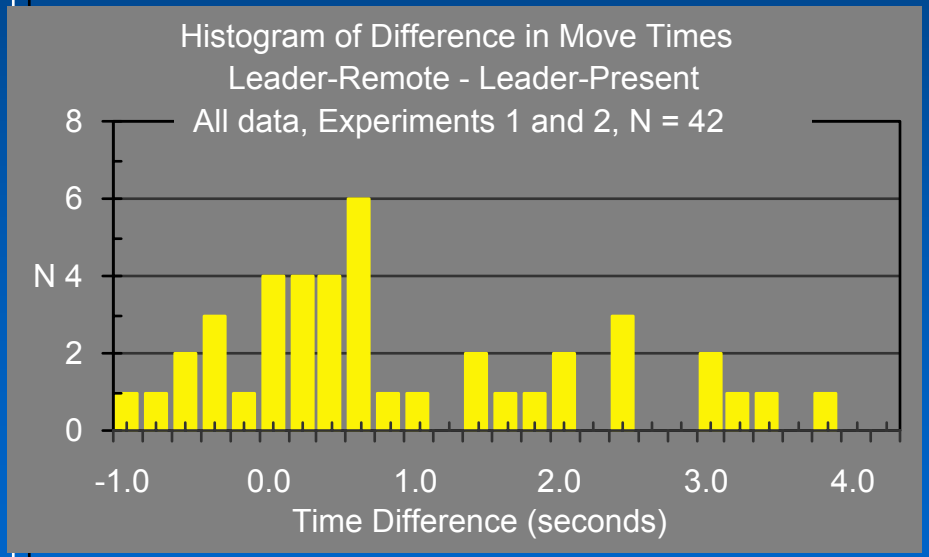
Participants: 22 students, 4 female, age 18 - 23.

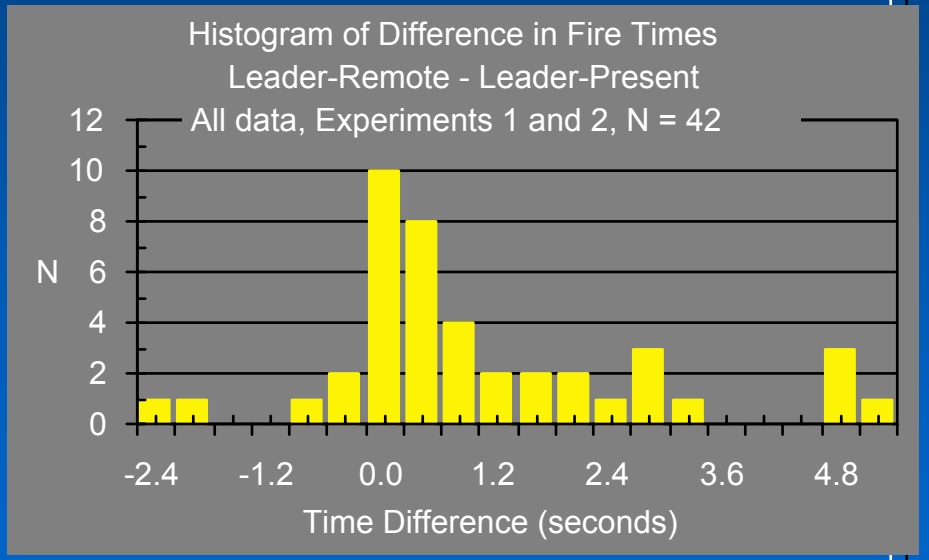
© Kip Smith, 2003

The paintball experiments

Results:

Response times to the commands to move and to shoot were significantly faster in the leader-present condition in both experiments. No significant differences were found across the two experiments. The graphs and statistical tests below aggregate data across experiments.





$t(41) = 4.122, p < .0001$

$t(41) = 3.218, p < .0020$

© Kip Smith, 2003

The paintball experiments 5

Implications:

Young adults are slower to obey commands given remotely than they are to commands given by leaders present in the field.

The paintball assault lane is a low-cost, natural laboratory for investigating the impact of leader presence on soldier performance and efficiency. It affords both application of experimental control and generalization to battlefield operations.

© Kip Smith, 2003

The paintball experiments

Proposed program, July - December 2003:

- Migrate the experimental paradigm to Sweden.
- Replicate experiments looking for cultural differences:
  - Individual mind (USA) vs. collective mind (Sweden).
- Enhance data collection:
  - Use telemetry to capture response times and psychophysiologic measures of stress.
- Extend concept from individuals to teams:
  - Conduct proof-of-concept studies of modified Army STX lanes to assess effect of leader presence on:
    - Trust in the leader by (5-person) team.
    - Trust among team-members.

© Kip Smith, 2003

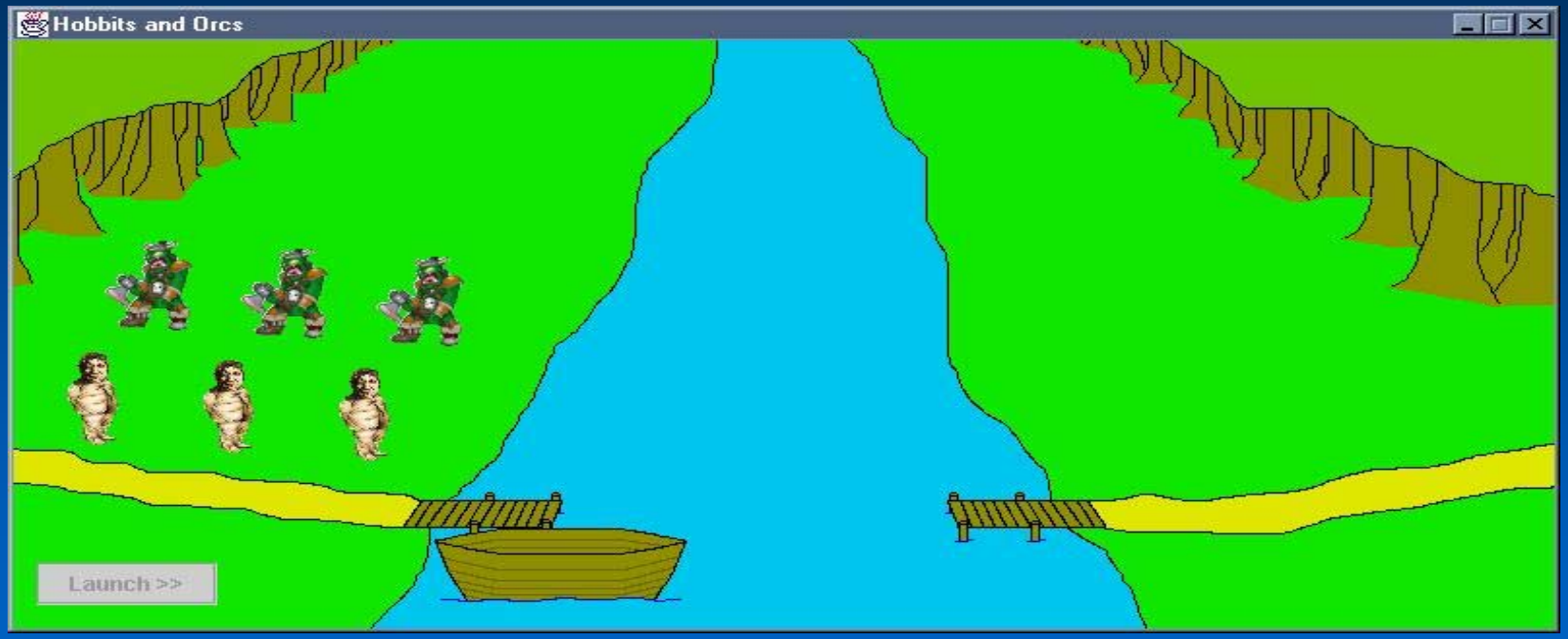
3B) Three experiments on the interaction of revealed trust, proximity, and communication mode on team cognitive performance

Experiment 1: Hobbits and Orcs
Experiment 2: The X Boat Game
Experiment 3: The Boat Game

© Kip Smith, 2003

Experiment 1: Hobbits and Orcs

On the impact of revealed trust on performance in a cooperative problem-solving task



© Kip Smith, 2003

Experiment 1: Hobbits and Orcs 1

On the impact of revealed trust on performance in a cooperative problem-solving task

Question:

How do revealed trust, proximity, and communication mode interact to affect team cognitive performance in a cooperative problem-solving task?

Hypotheses:

We expect cognitive performance to improve if:

A) Participants are permitted to communicate freely,
OR

B) They work face-to-face rather than over a radio link,
OR

C) They are both trusting and trustworthy.

© Kip Smith, 2003

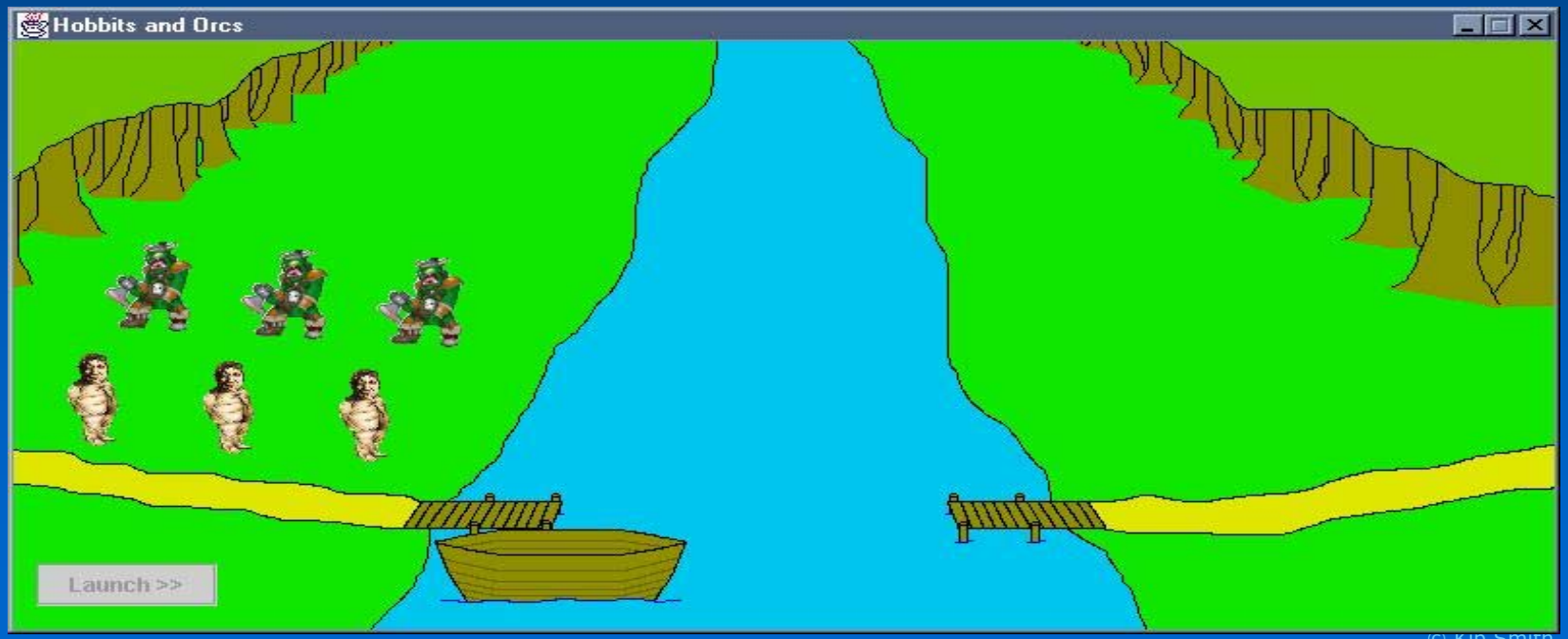
Experiment 1: Hobbits and Orcs

Task:

A distributed version of the classic Hobbits and Orcs river-crossing problem (Greeno, 1974).

Dependent measures:

Time on task, number of moves, number of illegal moves.



© Kip Smith, 2003

Experiment 1: Hobbits and Orcs 3

Experimental Design, a 2 X 2 X 3 Quasi-experiment

72 pairs of participants (18 female pairs, 23 male pairs, and 31 mixed pairs; median age 19, mode 18, maximum 48).

4 fully-crossed (2x2) between-sjs treatment levels with proximity at 2 levels and communication mode at 2 levels:

- colocated with free talk
- colocated with no talk
- remote with free talk using a 2-way radio
- remote no talk

Trust was a quasi- (independent) variable at three levels. Pairs in all 4 levels of proximity and communication mode played the extensive-form trust game to self-select into one of three categories of trust: distrusting, trusting/trustworthy, betraying.

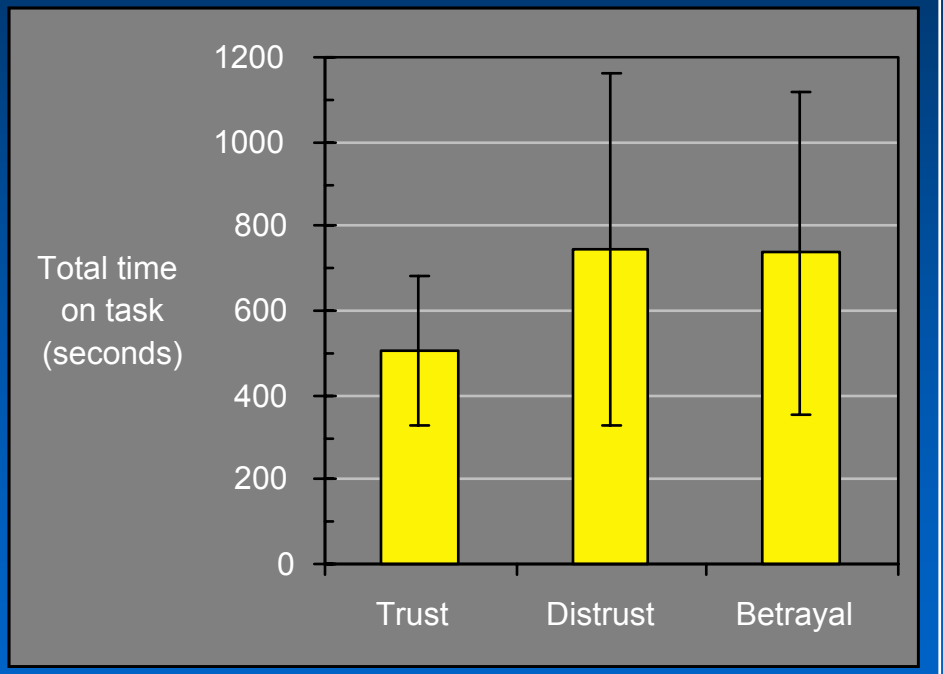
© Kip Smith, 2003

Experiment 1: Hobbits and Orcs 4

Results:

Both trust and communication mode had a significant impact on team performance. They did not interact. Proximity had no effect.

Teams that trusted needed less time to complete the task:



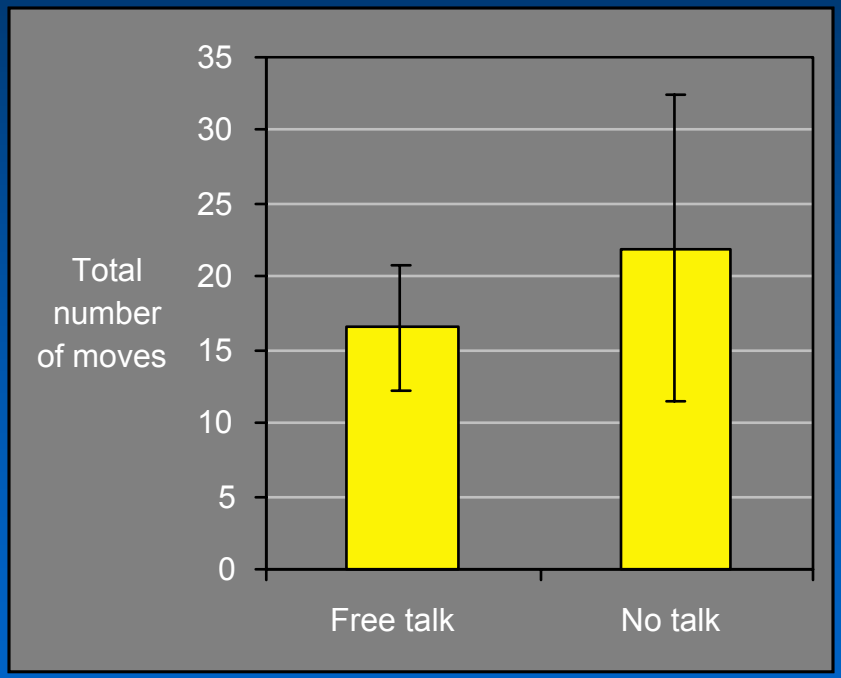
$F(2, 53) = 4.52, p < .016$

© Kip Smith, 2003

Experiment 1: Hobbits and Orcs

Results:

Teams that communicated freely (whether face-to-face or by radio) required fewer moves to complete the task:



$F(1, 53) = 8.69, p < .005$

Implications:

Trust improves effectiveness.

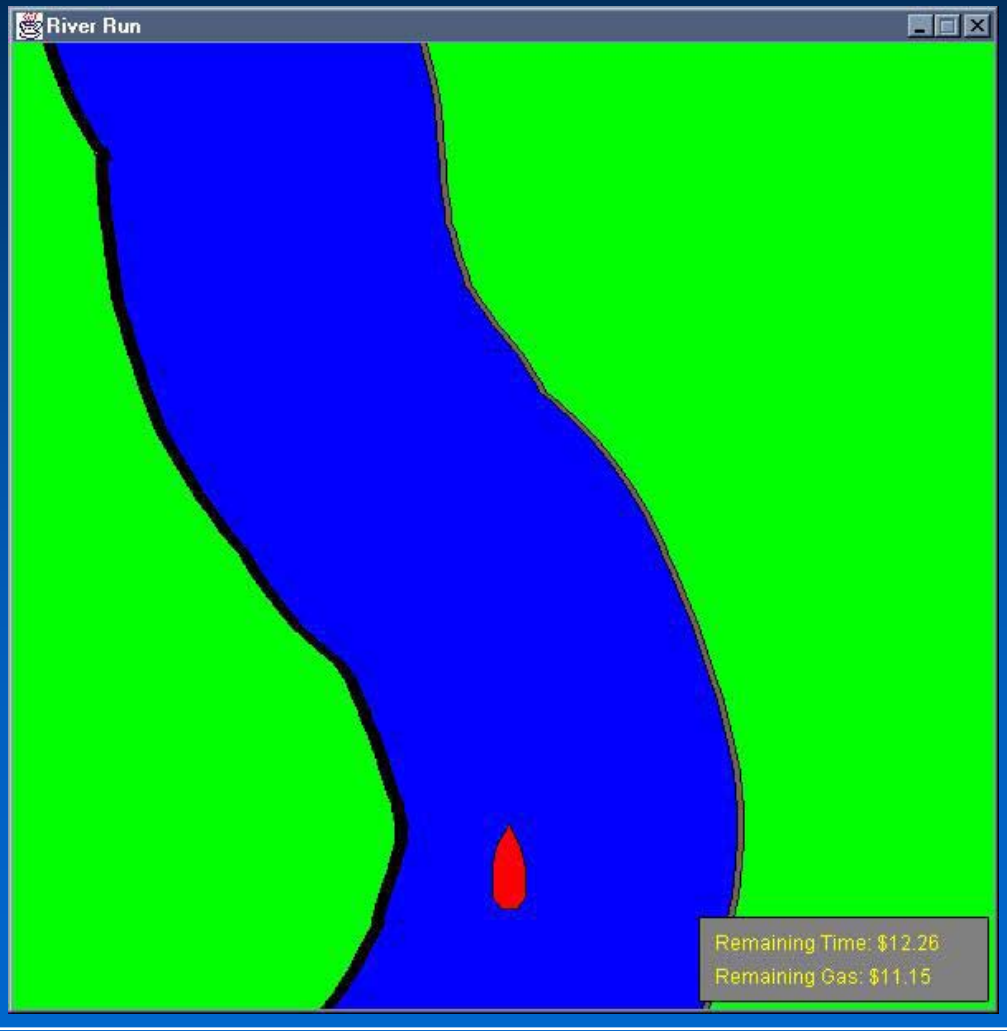
Communication improves team efficiency.

Radio communication does not degrade cooperative problem solving.

© Kip Smith, 2003

Experiment 2: X boats

On the impact of revealed trust on performance in a distributed command and control task



© Kip Smith, 2003

Experiment 2: X boats

On the impact of revealed trust on performance in a distributed command and control task

Question:

How do revealed trust, proximity, and communication mode interact to affect team cognitive performance in a distributed command and control task?

Hypotheses:

We expect cognitive performance to improve if:

A) Participants are permitted to communicate freely,
OR

B) They work face-to-face rather than over a radio link,
OR

C) They are both trusting and trustworthy.

© Kip Smith, 2003

Experiment 2: X Boats

Task:

Work together to navigate a tug boat up a sinuous river to the dock.

Player 1 is the pilot, responsible for steering.

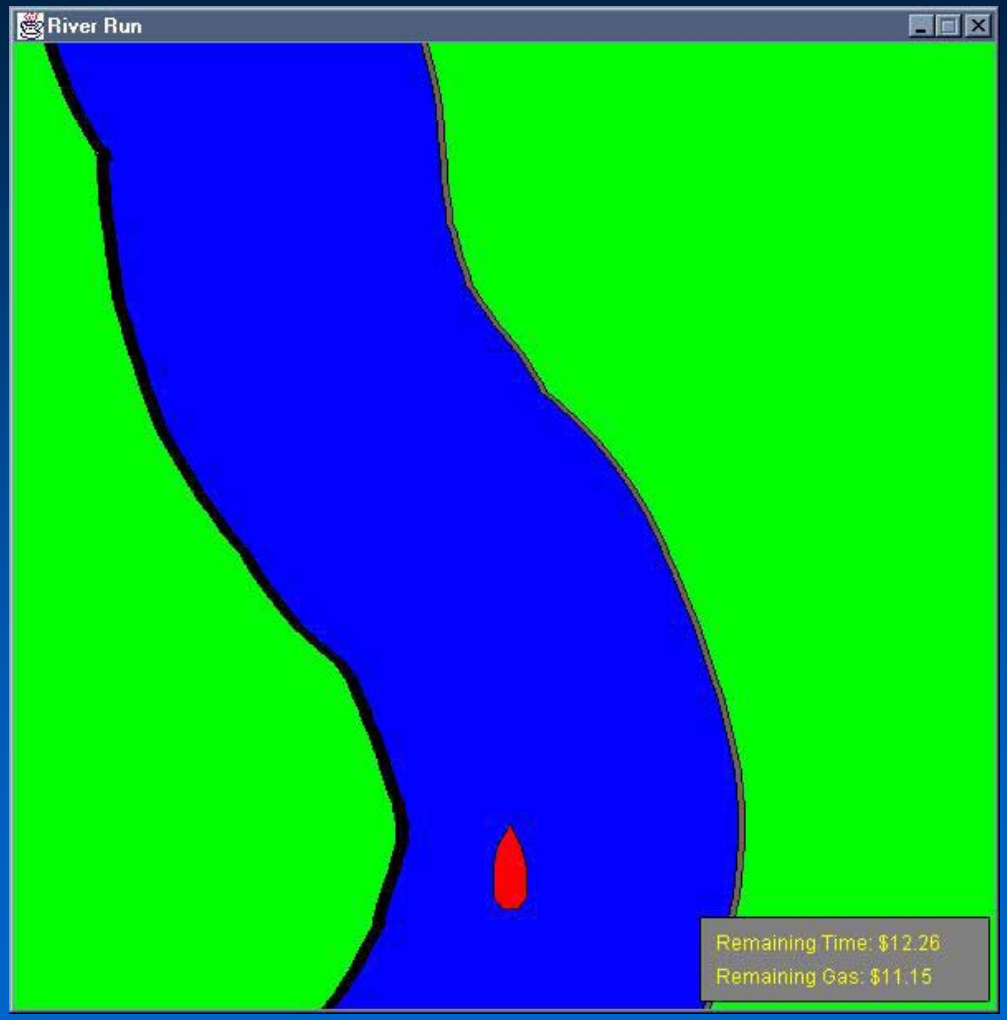
Player 2 is the engineer, responsible for speed.

Running into the bank is not good.

The boat has inertia and is slow to respond to all commands.

Dependent measures:

Time on task, number of errors (crashes into the bank).



© Kip Smith, 2003

Experiment 2: X boats

Experimental Design, a 2 X 2 X 3 Quasi-experiment

63 pairs of participants (19 female pairs, 17 male pairs, and 27 mixed pairs; median age 19, mode 18, maximum 28).

As in Experiment 1,

4 fully-crossed (2x2) between-sjs treatment levels with proximity at 2 levels and communication mode at 2 levels:

- colocated with free talk
- colocated with no talk
- remote with free talk using a 2-way radio
- remote no talk

Again, trust was a quasi- (independent) variable at three levels. Pairs in all 4 levels of proximity and communication mode played the extensive-form trust game to self-select into one of three categories of trust: distrusting, trusting/trustworthy, betraying.

© Kip Smith, 2003

Experiment 2: X Boats

Results:

40 teams trusted and were trustworthy.

17 teams experienced betrayal.

Only 6 teams experienced distrust, an insufficient number for statistical power. They have been excluded from the analysis.

We report the results of the 2x2x2 ANOVA on the remaining 57 teams where trusting and betraying are the two trust conditions.

© Kip Smith, 2003

Experiment 2: X Boats

Results:

Previously revealed trust interacts with both (1) proximity and (2) communication mode.

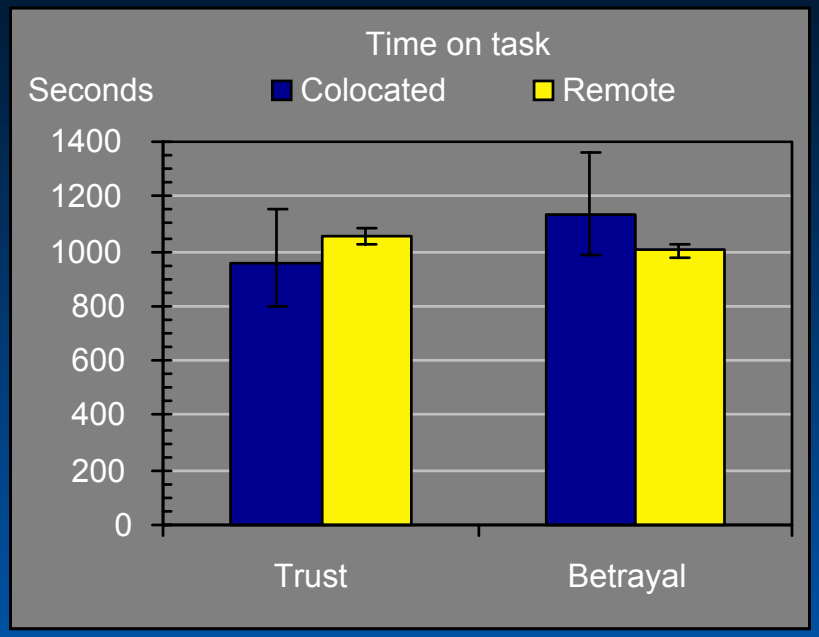
(1) Betrayal hampers efficiency when teams work face-to-face but not when they work remotely:

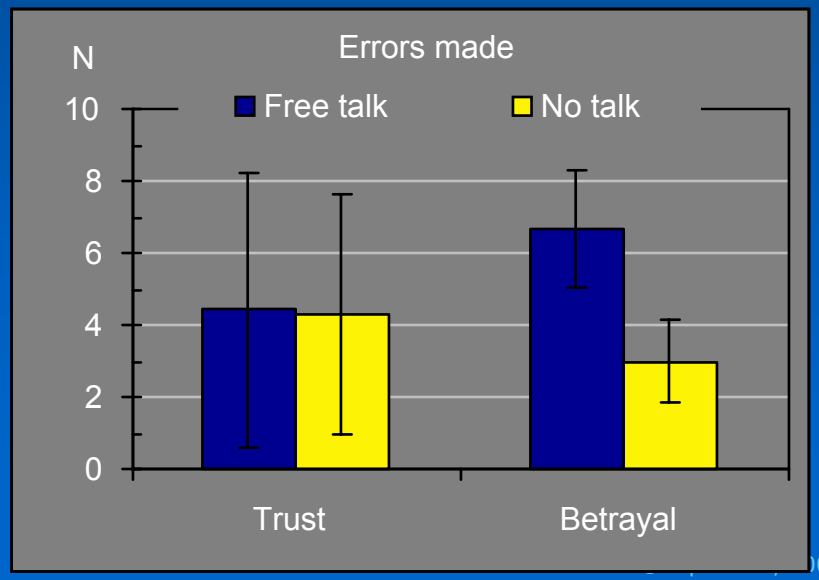
$F(1, 49) = 4.57, p < .038$

(2) Communication mode has no effect on error rate for teams that trusted and were trustworthy.

Teams that experienced betrayal were less effective if, and only if, they had to communicate:

$F(1, 49) = 3.49, p < .068$





Experiment 2: X Boats

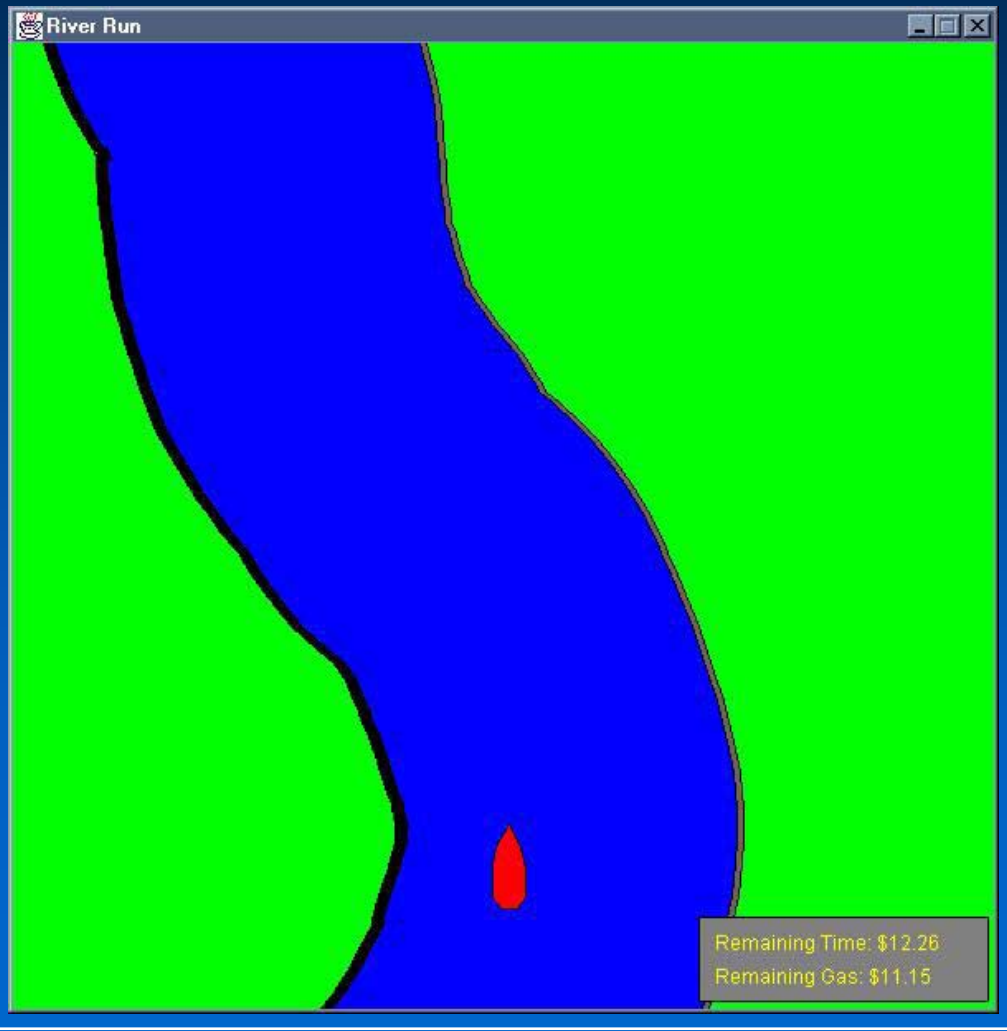
Implications:

Betrayal - the failure to reciprocate an expression of trust with trustworthiness - can have a profoundly negative influence on how well teams communicate and on how well they coordinate their activities when working face-to-face.

© Kip Smith, 2003

Experiment 3: The boat game

On the impact of performance in a distributed command and control task on revealed trust



© Kip Smith, 2003

Experiment 3: The boat game

On the impact of performance in a distributed command and control task on revealed trust

Question:

How do proximity, communication mode, and cognitive performance in a distributed command and control task interact to affect subsequent trusting behavior?

This experiment turns the question around to assess the direction of causation between trust and performance.

Hypotheses:

We expect the likelihood of trust to increase if:

A) Participants are permitted to communicate freely,
OR

B) They work face-to-face rather than over a radio link,
OR

C) They complete the task successfully.

© Kip Smith, 2003

Experiment 3: The boat game 2

2 X 2 Experimental design:

63 pairs of participants (19 female pairs, 17 male pairs, and 27 mixed pairs; median age 19, mode 18, maximum 40).

As in Experiments 1 and 2,
4 fully-crossed (2x2) between-sjs treatment levels with proximity at 2 levels and communication mode at 2 levels:

- colocated with free talk
- colocated with no talk
- remote with free talk using a 2-way radio
- remote no talk

However, in Experiment 3:
Trust is a dependent variable at three levels. After finishing the boat game, pairs played the extensive-form trust game to self-select into one of three categories of trust: distrust, trust/trustworthy, betrayal.

© Kip Smith, 2003

Experiment 3: The boat game

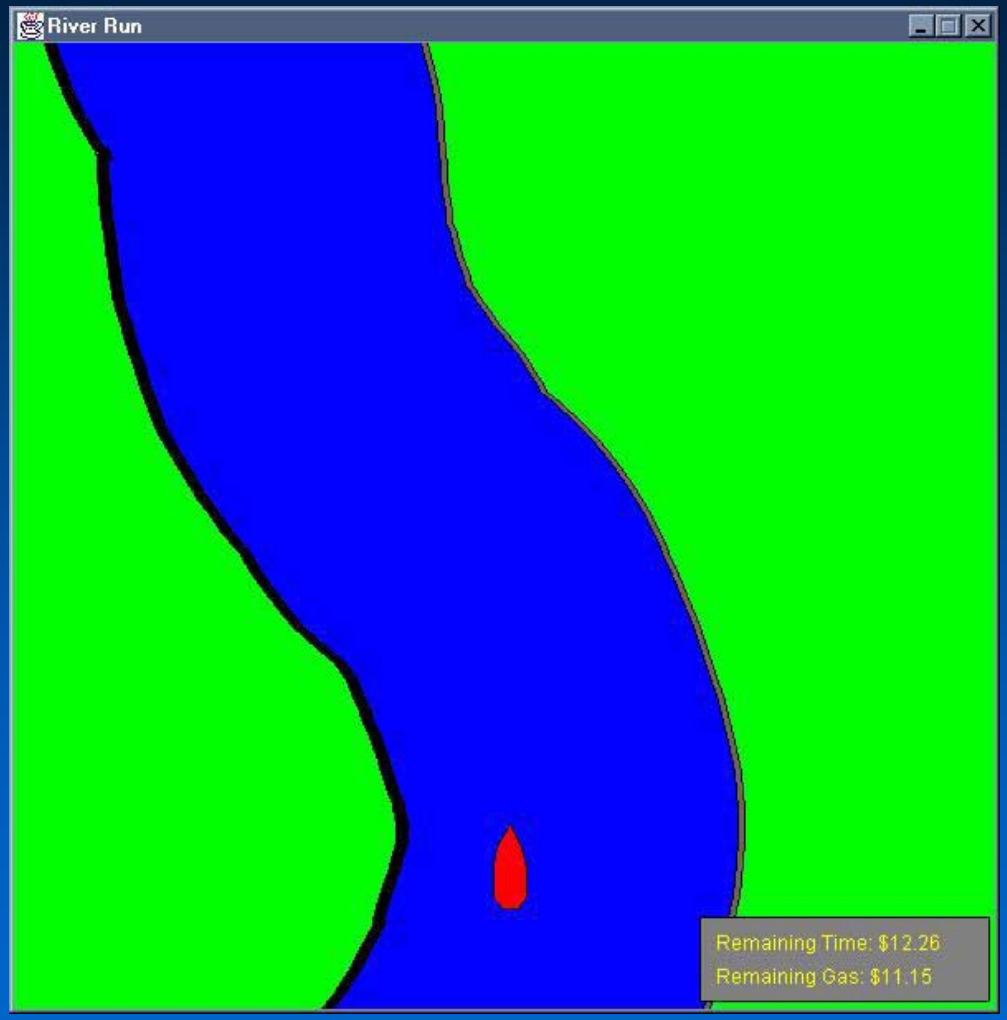
Task:

Same as in Experiment 2.

- Work together to navigate a tug boat up a sinuous river to the dock.
- Player 1 is the pilot, responsible for steering.
- Player 2 is the engineer, responsible for speed.
- Running into the bank is not good.
- The boat has inertia and is slow to respond to all commands.

Dependent measures:

Time on task, number of errors (crashes into the bank), & revealed trust.



© Kip Smith, 2003

Experiment 3: The boat game

Results:

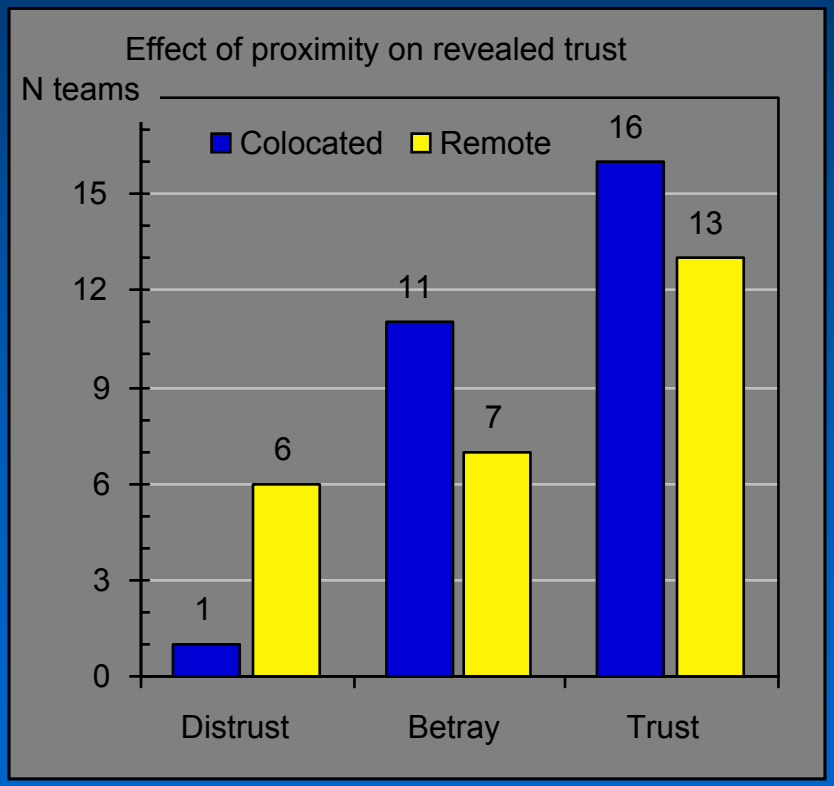
As in Experiment 2, communication mode and proximity had no effect on errors made and time on task.

In this experiment, trust is a categorical dependent variable.

Logistic regressions reveal:

- Quality of performance has no effect on revealed trust.
- Proximity has a strong effect on revealed trust:
  - Teams that worked face-to-face were more likely to trust and less likely to distrust.

$\chi^2(2) = 1.31, p < .02$



© Kip Smith, 2003

Experiment 3: The boat game

Implications:

Proximity played a critical role in the formation of trust between members of our teams. Teams that worked together were more likely to trust and be trustworthy.

When a combat team is initially assembled, it would be wise to bring all members physically together for an engaging task to mitigate the potential emergence of underlying distrust.

© Kip Smith, 2003

Student Collaborators

Shala Bloomberg
Jessica Bolton
Nathan Fosse
Steve Freund
Whitney Gee
Jill Halleran
Serena Hawkins
Andy Kearney
Holly Pangburn
2Lt Keith Pangburn
Nicole Pruitt
Andrea Shane
Jana Smith
Julie Soupene
Ben Swan
Molly Toll
Beth Younger

© Kip Smith, 2003

Acknowledgements

This work was supported by the DOD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Dr. Elmar Schmeisser is the project monitor. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.



© Kip Smith, 2003

MURI OPUS Research at Linköping University (LiU), Sweden

Third Quarter 2003 Progress Report

Revealed trust as a source of stress on the electronic battlefield

Kip Smith, Ph.D.
Principal Investigator

Peter Hancock
MURI OPUS Principal Investigator

Contents

- Two column goal statement

- Completed and ongoing work
  - Moving to Sweden
  - The paintball experiments: Leader presence
  - Laboratory experiments: Revealed trust and team cognitive performance
- Future plans
  - Fire teams and leader presence
  - Cultural diversity and trust in distributed command and control

Two Column Goal Statement

- 3rd quarter goals met

✓ Move the K State subaward and research to Linköping University, LiU.

✓ Lab experiments
  - Expand the lab experiments using the economic trust game at LiU.
  - Begin to collect psychophysiologic data for assessments of stress.

✓ Collaboration
  - Establish research links with the Swedish military and other institutions.
  - Establish links with ARI and ARL to inform the design of paintball experiments.

- Goals for 2003-2004

- Paintball experiments
  - Expand the paintball paradigm from soldier-leader to fire-team-leader using Army training scenarios from Ft Benning.
  - Test the ARL framework for understanding cultural diversity in cognition and teamwork.

- Lab experiments
  - Test hypotheses about the centipede game as a tool for establishing common ground and instilling trust.

Moving to Linköpings universitetet

- All personnel and equipment survived the move intact and without severe system failures.
  - It is more difficult to move $$ than to move people and equipment.

- Approvals to run experiments with human subjects have been obtained.
  - The Swedish IRB is vastly different than the American.
  - ➔ The role of culture in decision making.

LiU Collaboration

- Försvarshögskolanet
  - Prof. Berndt Brehmer, Professor of Command and Control Science at the Swedish National Defense College in Stockholm.
  - On the impact of trust and distrust on team performance in distributed warfare centers.
- ARI
  - Dr. Don Lampton, Orlando, FL
  - Dr. Robert Pleban, Fort Benning, GA.
  - Scenarios for studies of leader presence under live fire (paintball).
- ARL
  - Dr. Janet Sutton, Fort Sill, OK.
  - Framework for studies of cultural diversity in distributed command and control.

Summaries of LiU MURI OPUS research program

- 1) The effect of leader presence and communication mode on soldier performance under live fire.

- 2) Three laboratory studies on the interaction of revealed trust, proximity, and communication mode on team cognitive performance.

The effect of leader presence and communication mode on soldier performance under live fire

The Paintball Experiments

Experiment 1: Range 52, Fort Riley
Experiment 2: Quinlan Gardens, KSU

The paintball experiments

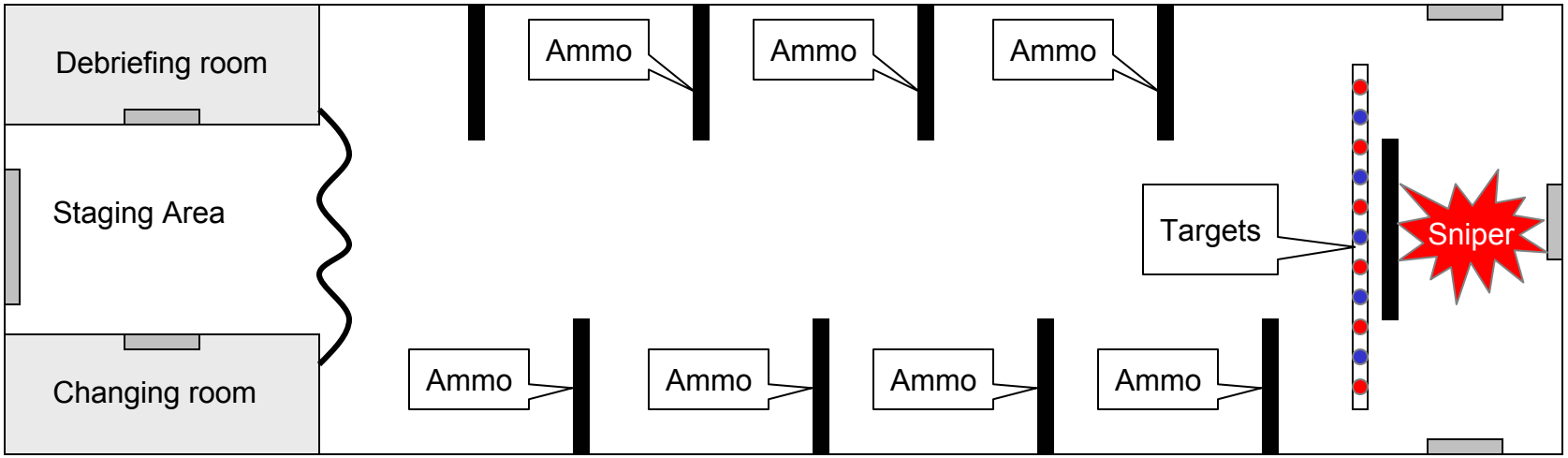
- Question:
  - Does leader presence on the battlefield influence the efficiency and effectiveness of a combat unit?
- Hypothesis:
  - The physical proximity of a commanding officer speeds the soldier's compliance with a command.
- Within-subjects design
  - Control condition: Leader in the assault lane with the soldier, communicating face-to-face.
  - Experimental condition: Leader out-of-sight, communicating by radio.
  - Order counterbalanced across participants.

The paintball experiments

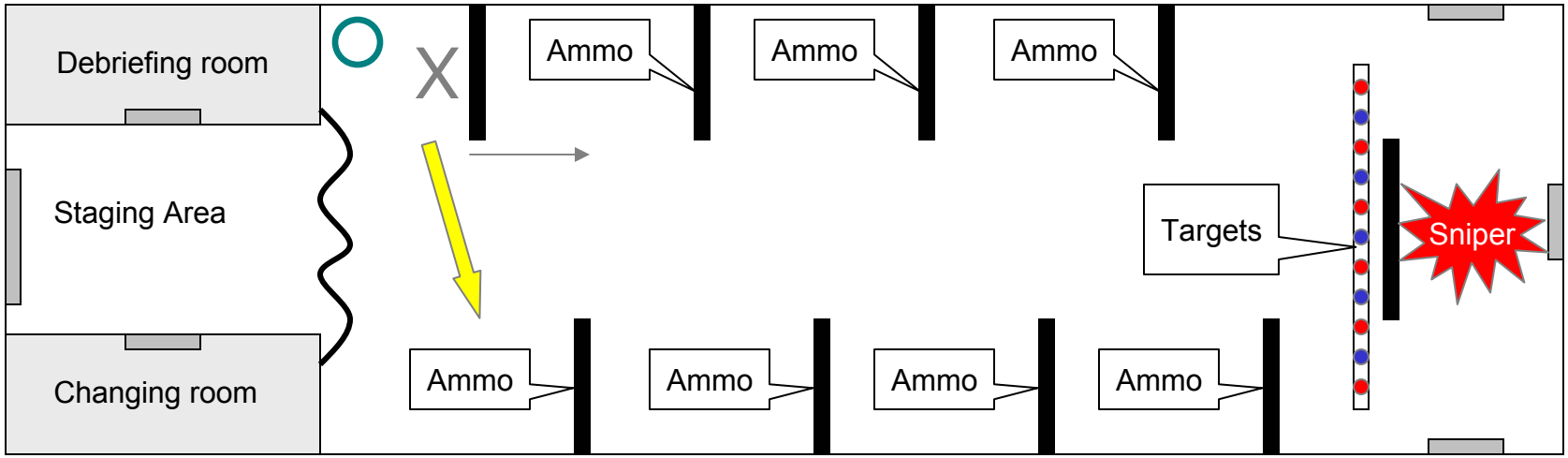
- Task:
  - Shoot, move, and communicate with the leader while under fire from a sniper shooting paintballs. Move up the assault lane to capture flags at the sniper's hardpoint.

- Dependent variables:
  - Soldier's response times to the leader's commands to shoot and to move up the lane.

The paintball assault lane



Procedure

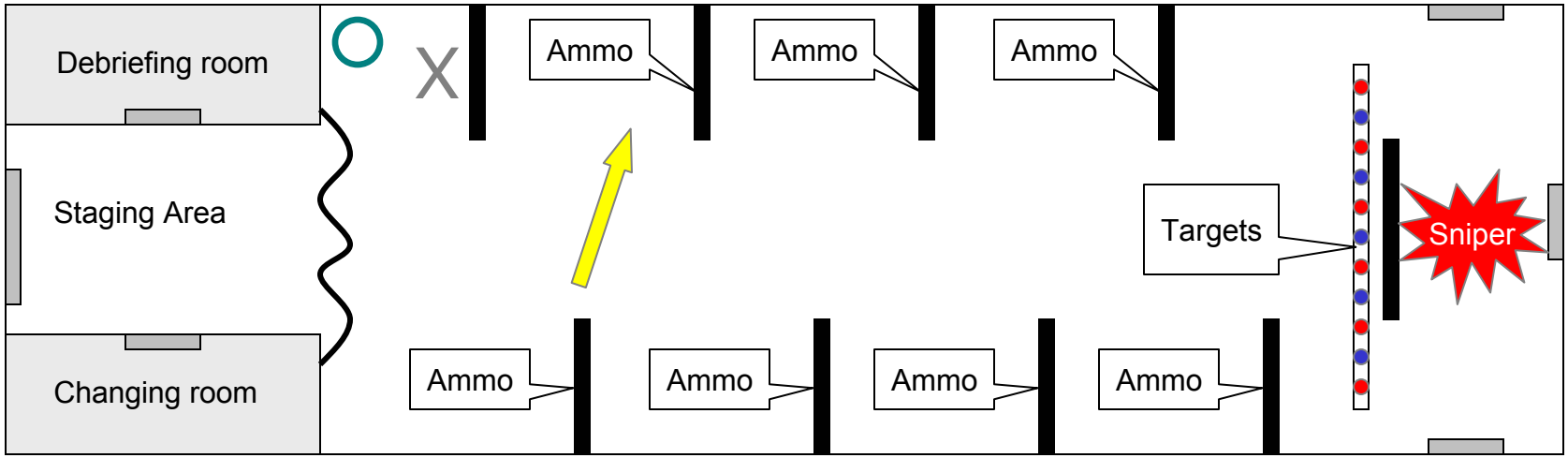


"Fire"

"Out of Ammo"

"Move"

Procedure



"Loaded"

"Fire"

"Out of Ammo"

"Move"

Experiment 1:
Range 52, Fort Riley, Kansas.

- Indoor lane
  - In the leader-remote condition, the leader was outdoors, completely out of the participant's sight.
- Participants
  - 20 students, 2 female, age 18 - 27.

Move times, Experiment 1

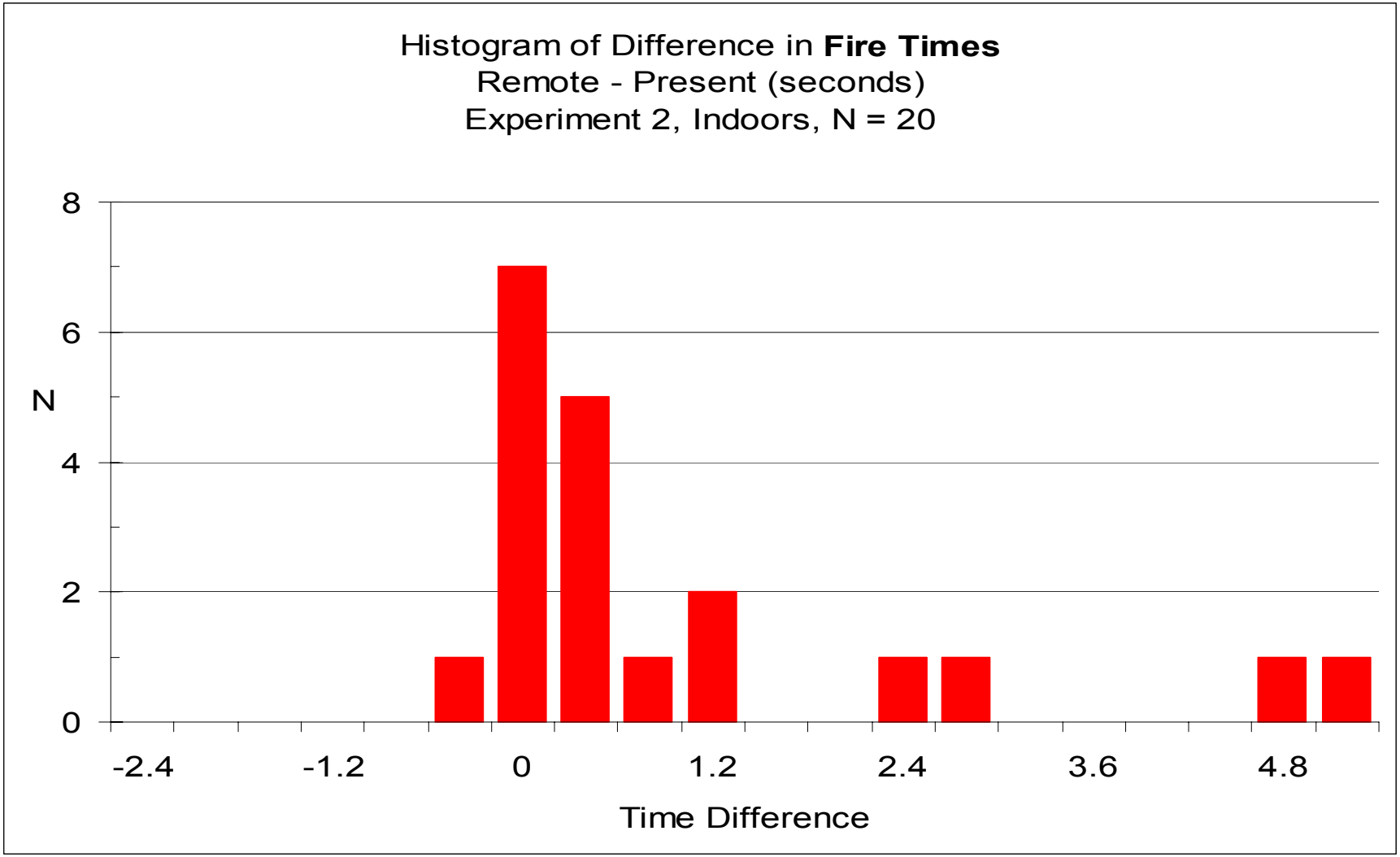


Move times, Experiment 1

- Of the 20 participants, 15 took longer to move in the Leader-Remote condition than the Leader-Present condition.

- $t(19) = 2.958$, $p < .004$

Fire times, Experiment 1



Fire times, Experiment 1

- Of the 20 participants, 12 took longer to fire in the leader-remote condition than the leader-present condition.

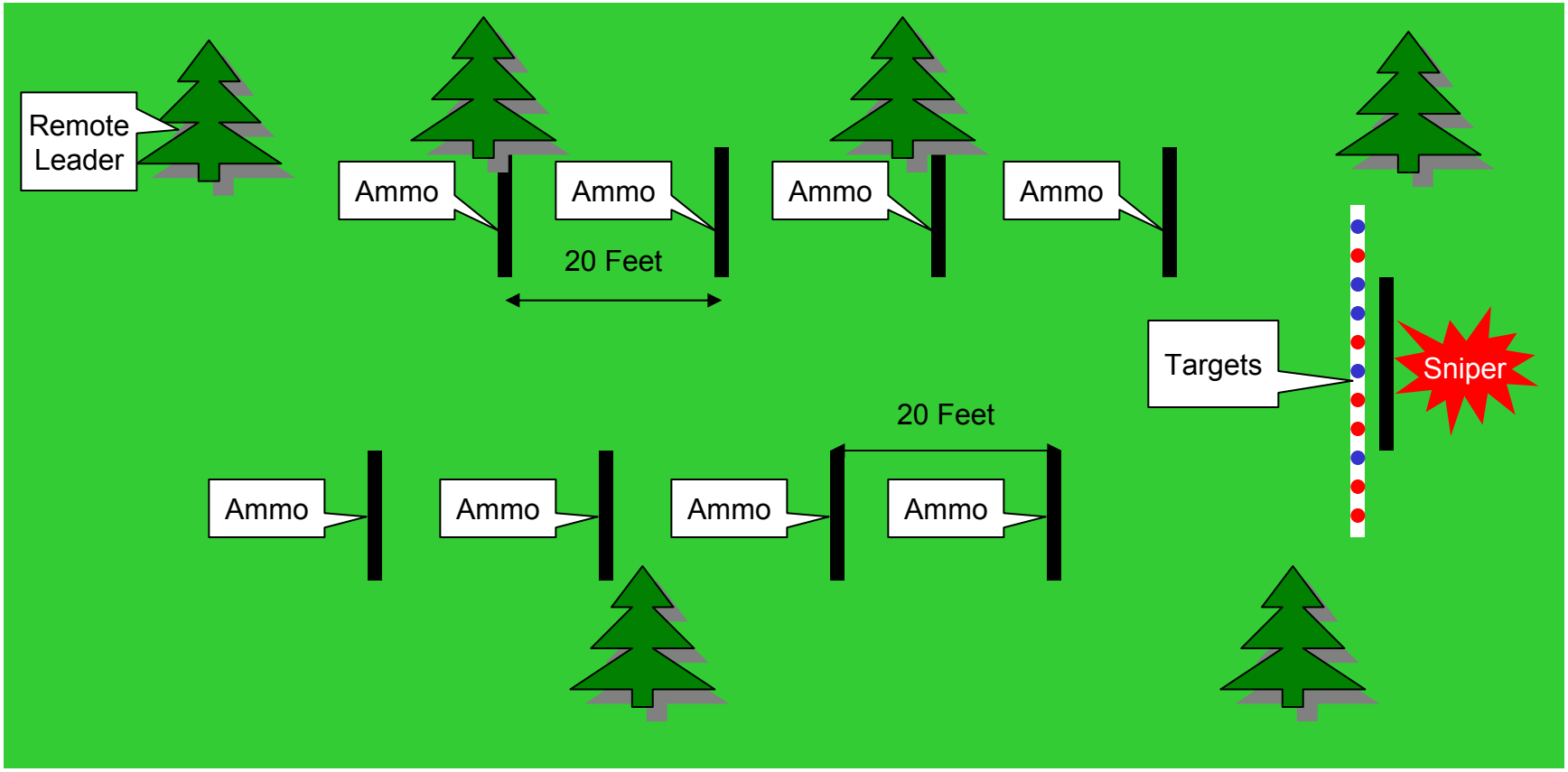
- $t(19) = 2.317$, $p < .016$

Experiment 2:
Quinlan Gardens, K State campus.

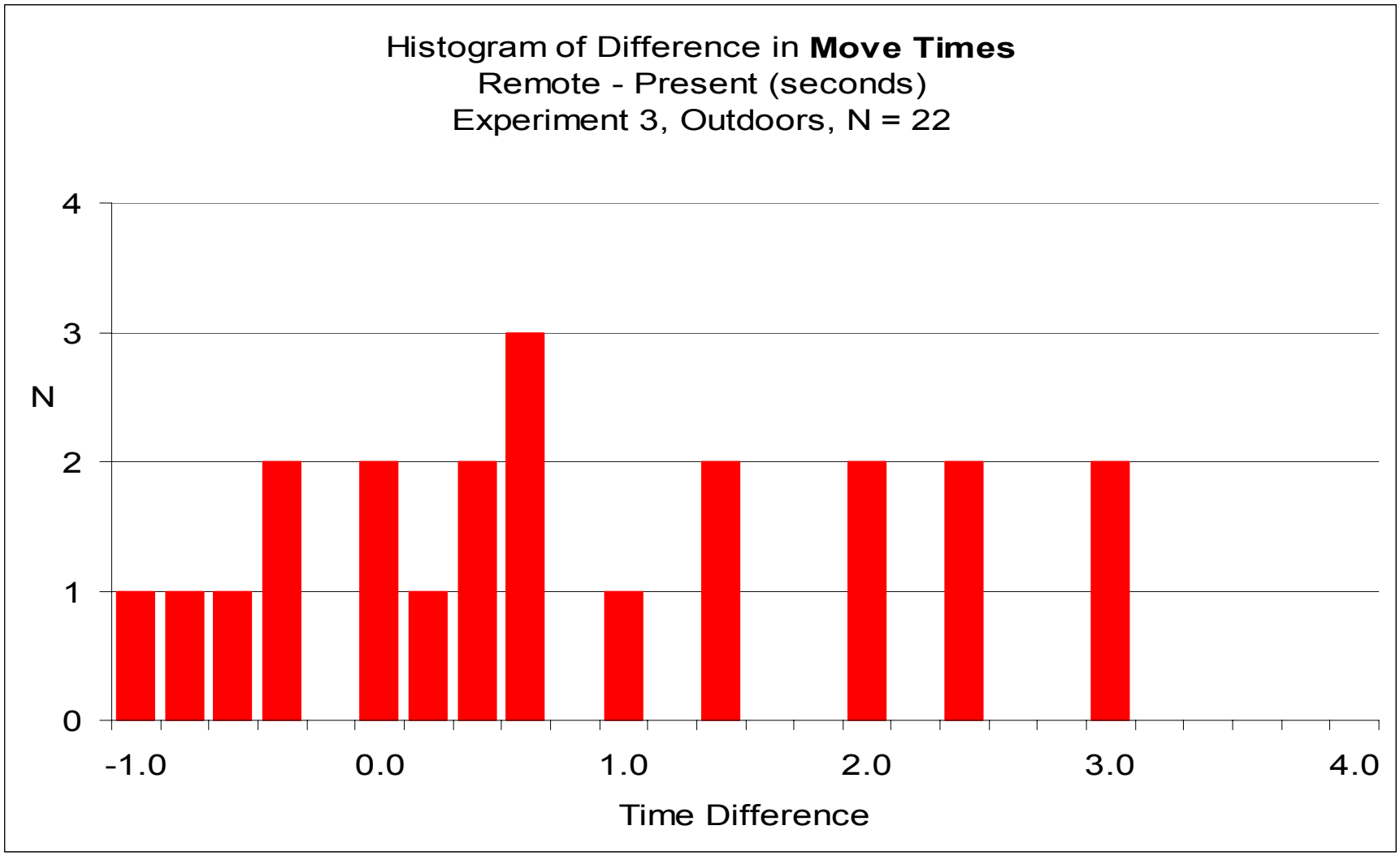
- Outdoor lane
  - In the leader-remote condition, the leader was ± 20 m behind the lane, out of the participant's line of sight but visible if the participant were to turn around and look.

- Participants
  - 22 students, 4 female, age 18 - 23.

Assault Lane, Experiment 2



Move times, Experiment 2

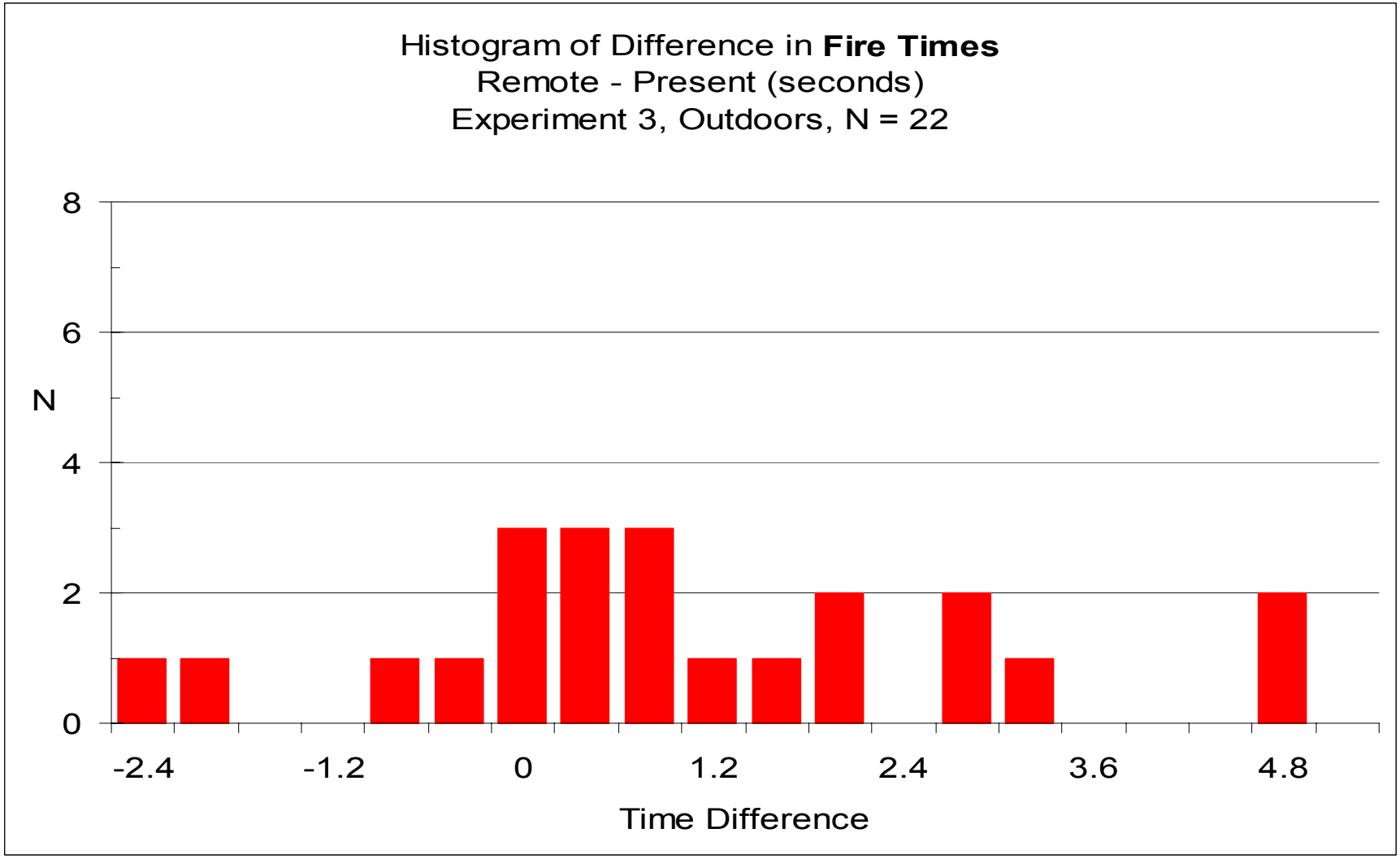


Move times, Experiment 2

- Of the 22 participants, 16 took longer to move in the leader-remote condition than the leader-present condition.

- $t(21) = 2.798$, $p < .006$

Fire times, Experiment 2



Fire times, Experiment 2

- Of the 22 participants, 15 took longer to fire in the leader-remote condition than the leader-present condition.

- $t(21) = 2.211$, $p < .020$

Composite results

- Move Times
  - $t(41) = 4.122$, $p < .0001$
- Fire Times
  - $t(41) = 3.218$, $p < .0013$

Implications for remote command and control

- Untrained but motivated young adults were slower to obey commands given remotely than they were to commands given by a leader present in the field.

- Remote command and control compromised the participants' response to commands to move and to fire.

- The degree of leader remoteness did not moderate this effect.

On the paintball paradigm

- The paintball assault lane is a low-cost, natural laboratory for investigating the impact of leader presence on soldier performance and efficiency.

- It affords both application of experimental control and generalization to battlefield operations.

Current experiments

- Recreating the study of leader presence with Swedish leaders and participants.

- Electronically capturing response times
  - Goniometers

- Telemetry of psychophysiologic data
  - Heart rate
  - Respiration
  - Galvanic skin response

Planned activity: Paintball, short term

- Expand study of leader presence (remote leader vs. present leader) on commands to move and to fire
- From soldier-leader

- To fire-team-leader
- Using scenarios currently used in virtual and field training at Fort Benning.
    - ARI contact: Dr. Robert J. Pleban

Planned activity: Paintball, long term

- Conduct cross-cultural studies to test ARL framework for understanding cultural diversity in cognition and teamwork
  - ARL contact: Dr. Janet L. Sutton, Fort Sill

- Swedes
  - Linköping
- Americans
  - Central Florida

ARL framework for cultural diversity in cognition and teamwork

National Cultural Dimension	Range	Taxonomy of tasks: Situation assessment	Coordination	Assigning roles and responsibilities	Support behavior
Acceptability of power differentials	High				
	Low				
Need for certainty	High				
	Low				
Activity orientation	Independent				
	Inter-dependent				

Modified from:

Sutton, J. L., & Pierce, L. G. (2003). A framework for understanding cultural diversity in cognition and teamwork. Proceedings of the 8th International Command and Control Research and Technology Symposium.

American culture

National Cultural Dimension	Americans Range	Taxonomy of tasks Situation assessment	Coordination	Assigning roles and responsibilities	Support behavior
Acceptability of power differentials	High				
	Low				
Need for certainty	High				
	Low				
Activity orientation	Independent				
	Inter-dependent				

Swedish culture

National Cultural Dimension	Swedes Range	Taxonomy of tasks Situation assessment	Coordination	Assigning roles and responsibilities	Support behavior
Acceptability of power differentials	High				
	Low				
Need for certainty	High				
	Low				
Activity orientation	Independent				
	Inter-dependent				

Three laboratory studies on the interaction of revealed trust, proximity, and communication mode on team cognitive performance

Experiment 1: Trust ➔ Cooperative decision making

Experiment 2: Trust ➔ Distributed command and control

Experiment 3: Distributed C^2 ➔ Trust

Issues

- What is the role of trust in distributed command and control?
  - Does trust influence team performance ?
  - Does team performance influence trust ?
  - What are the effects / interactions of trust with proximity and communication?

Outline

- Trust ➔ Performance ? / Performance ➔ Trust ?
- Hypotheses
  - Dissociation of influence of trust on individuals and teams
- Three experiments
  - Method
  - Results
- Implications
  - On the utility of the extensive form game
  - The U.S. Army's vision of the Objective Force Warrior

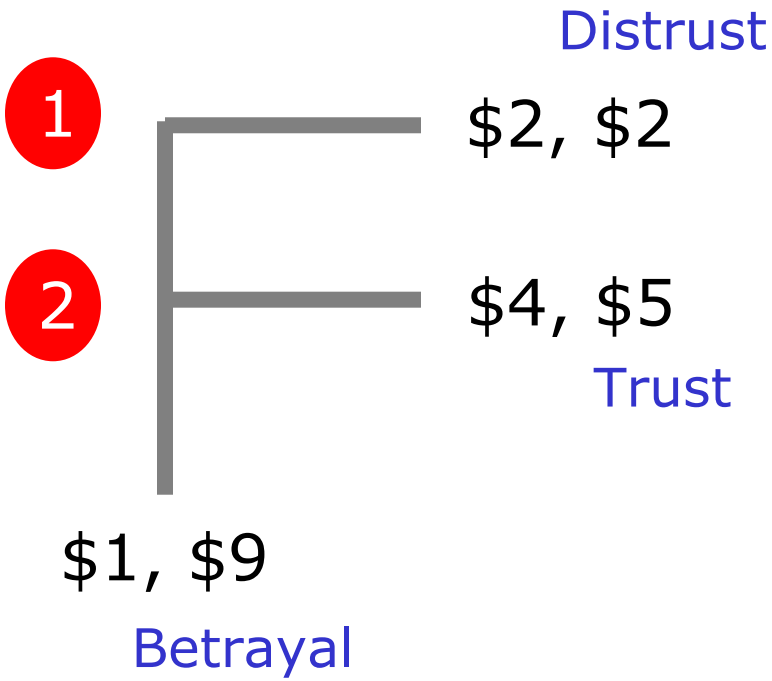
Assumptions

- We assume that trust is revealed by behavior.
- There are four faces of trusting behavior.
  - The trusting party may exhibit either
    - Distrust or
    - Trust
  - The trusted party may exhibit either
    - Trustworthiness or
    - Betrayal

Assumptions

- Trust behavior is fully and accurately revealed by play in the extensive form trust game (the F game).
  - McCabe, Rassenti, & V. Smith, 1995

In our experiments, the moves made by participants reveal the team's level of trust.



Hypotheses

- Revealed trust will influence subsequent team performance in cooperative, distributed tasks.
  - Trusting/trustworthy teams will perform better than distrusting teams and betraying teams.

- Revealed trust will not influence individual performance on simple cognitive tasks.

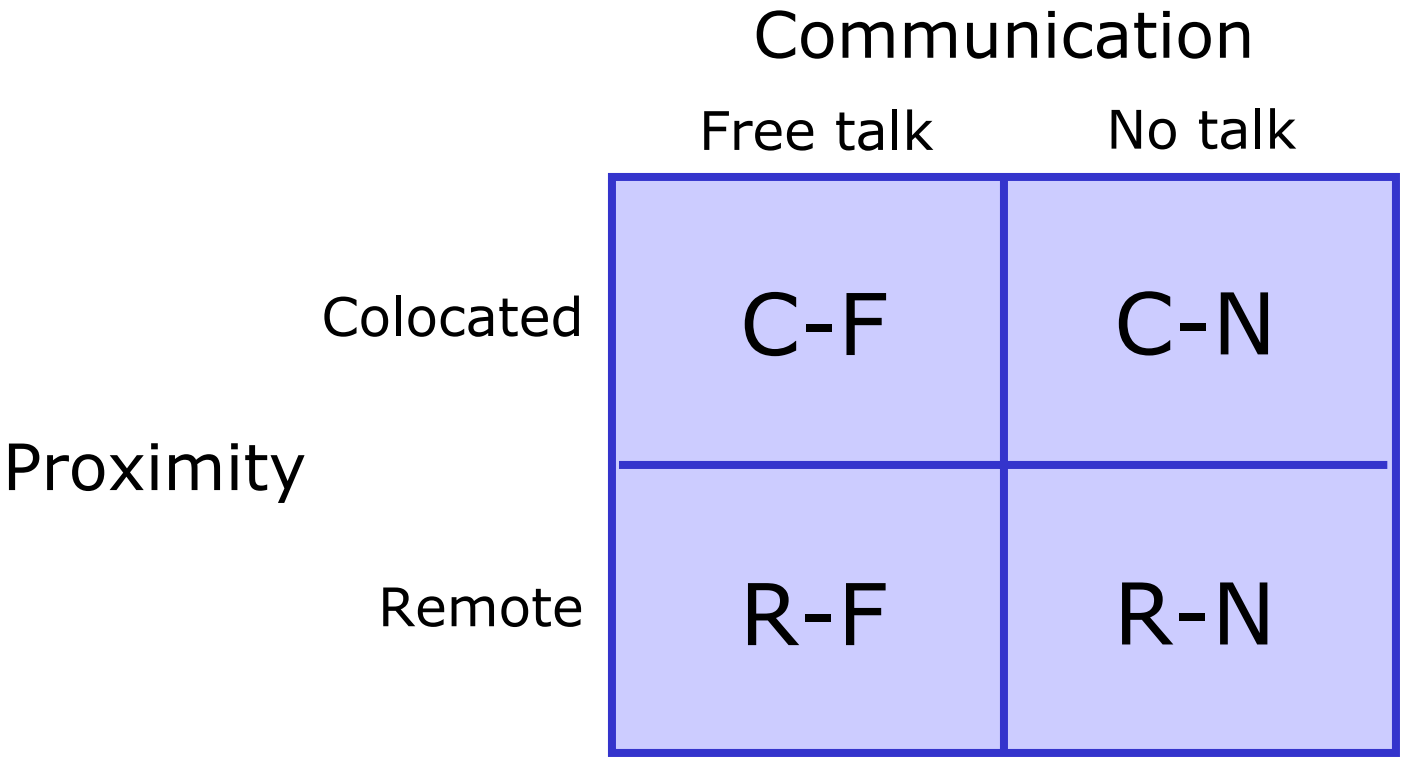
- Proximity influences team performance.
  - Colocated teams will perform better than distributed teams.
- Mode of communication influences team performance.
  - Teams that are allowed to talk will perform better.

Method, Experiments 1 and 2

- A series of four tasks
  - Individual cognitive task
  - Quasi-manipulation of trust using the F game
  - Team task
  - Repeated play of individual task

Experimental Design

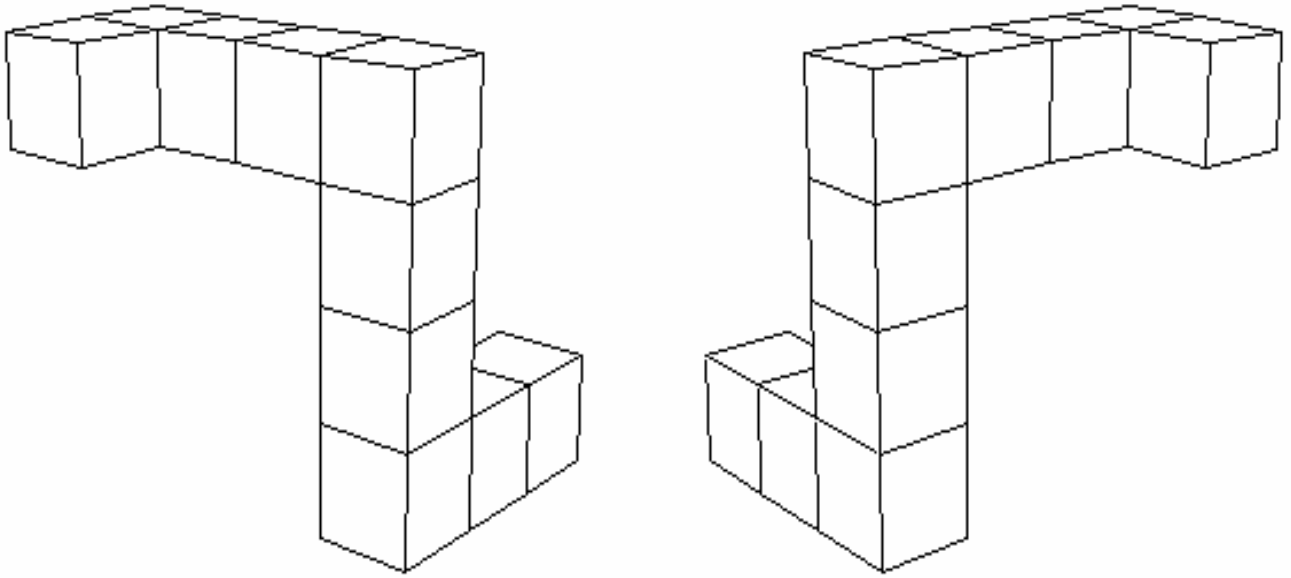
- 2 x 2 x 3 between subjects (teams) design
  - Proximity x Communication x Revealed trust



Experiment 1: F game ➔ Hobbits and Orcs

On the impact of revealed trust on performance in a cooperative problem-solving task

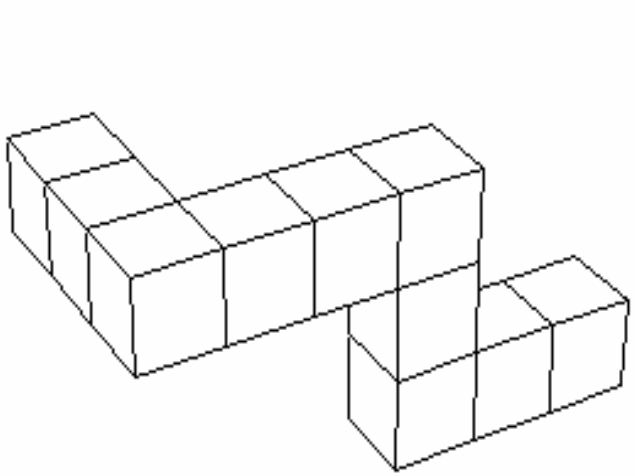
Individual task: Mental rotation

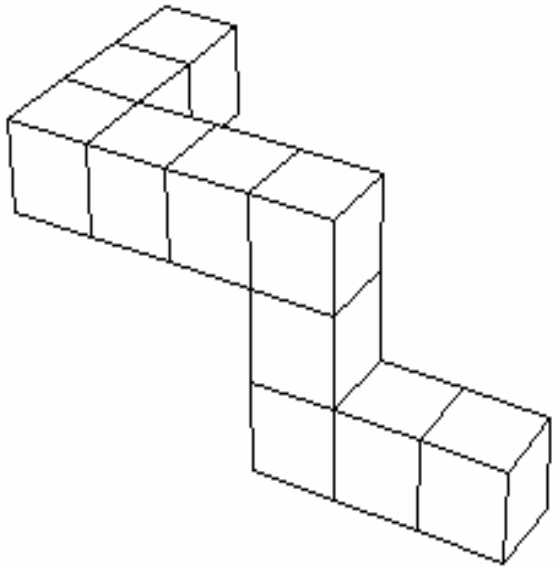


Are these figures the same or different?

Different

Mental rotation





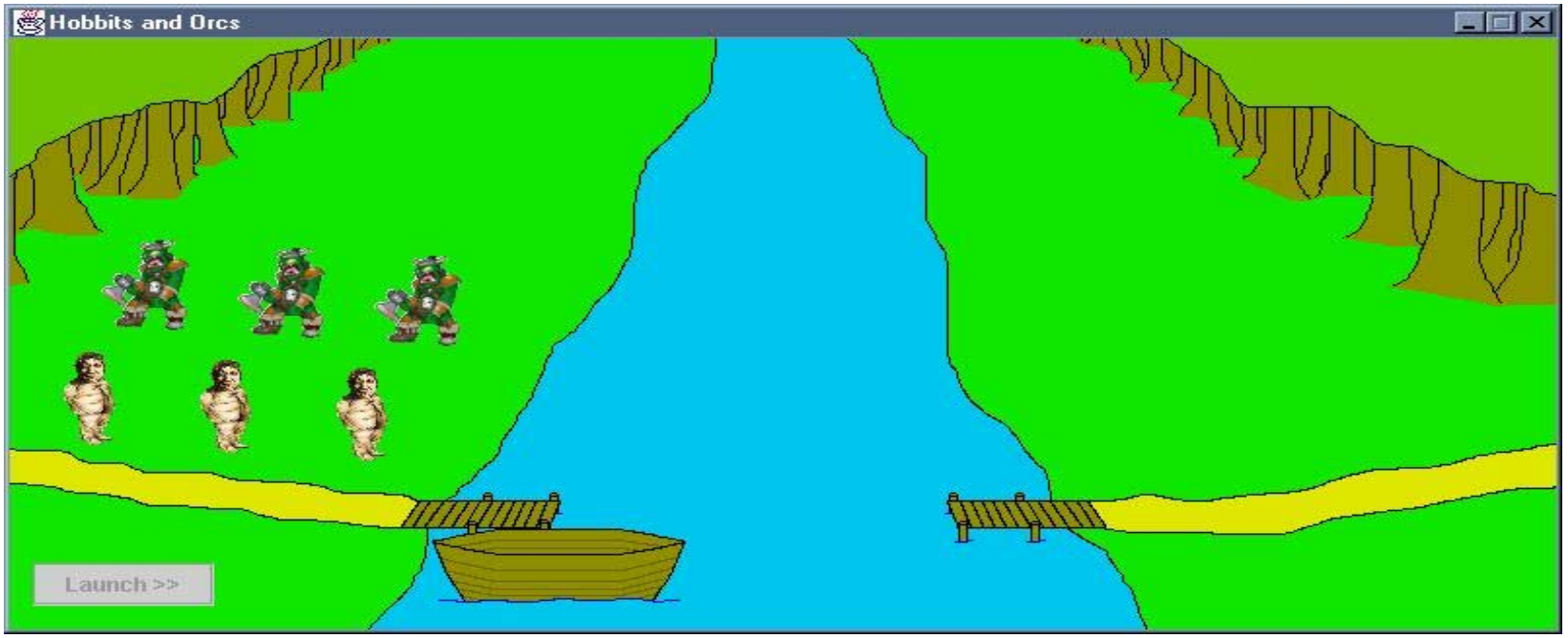
Are these figures the same or different?

Same

Team task: Hobbits and Orcs

- A distributed version of the classic Hobbits and Orcs river-crossing problem (Greeno, 1974).

- Dependent measures:
  - Time on task, number of moves, number of illegal moves.



Experiment 1: Hobbits and Orcs

- Question:
  - How do revealed trust, proximity, and communication mode interact to affect team cognitive performance in a cooperative problem-solving task?

- Hypotheses:
  - We expect cognitive performance to improve if:
  - A) Participants are permitted to communicate freely,
  - OR
  - B) They work face-to-face rather than over a radio link,
  - OR
  - C) They are both trusting and trustworthy.

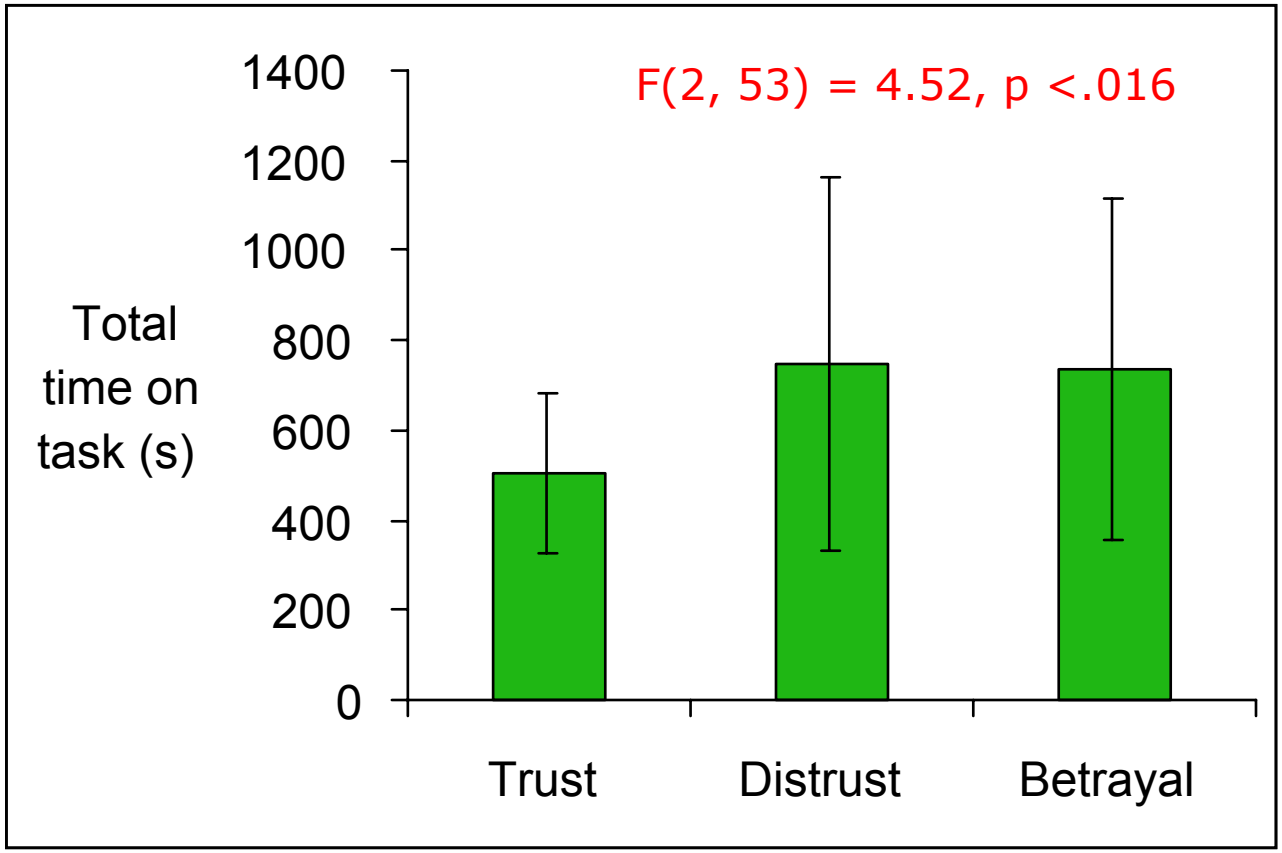
Participants, procedure

- 72 pairs of participants (18 female pairs, 23 male pairs, and 31 mixed pairs; median age 19, mode 18, maximum 48).

- Trust was a quasi- (independent) variable at three levels.
  - Pairs in all 4 levels of proximity and communication mode played the F game to self-select into one of three categories of trust: distrusting, trusting/trustworthy, betraying prior to working through the Hobbits and Orcs problem.

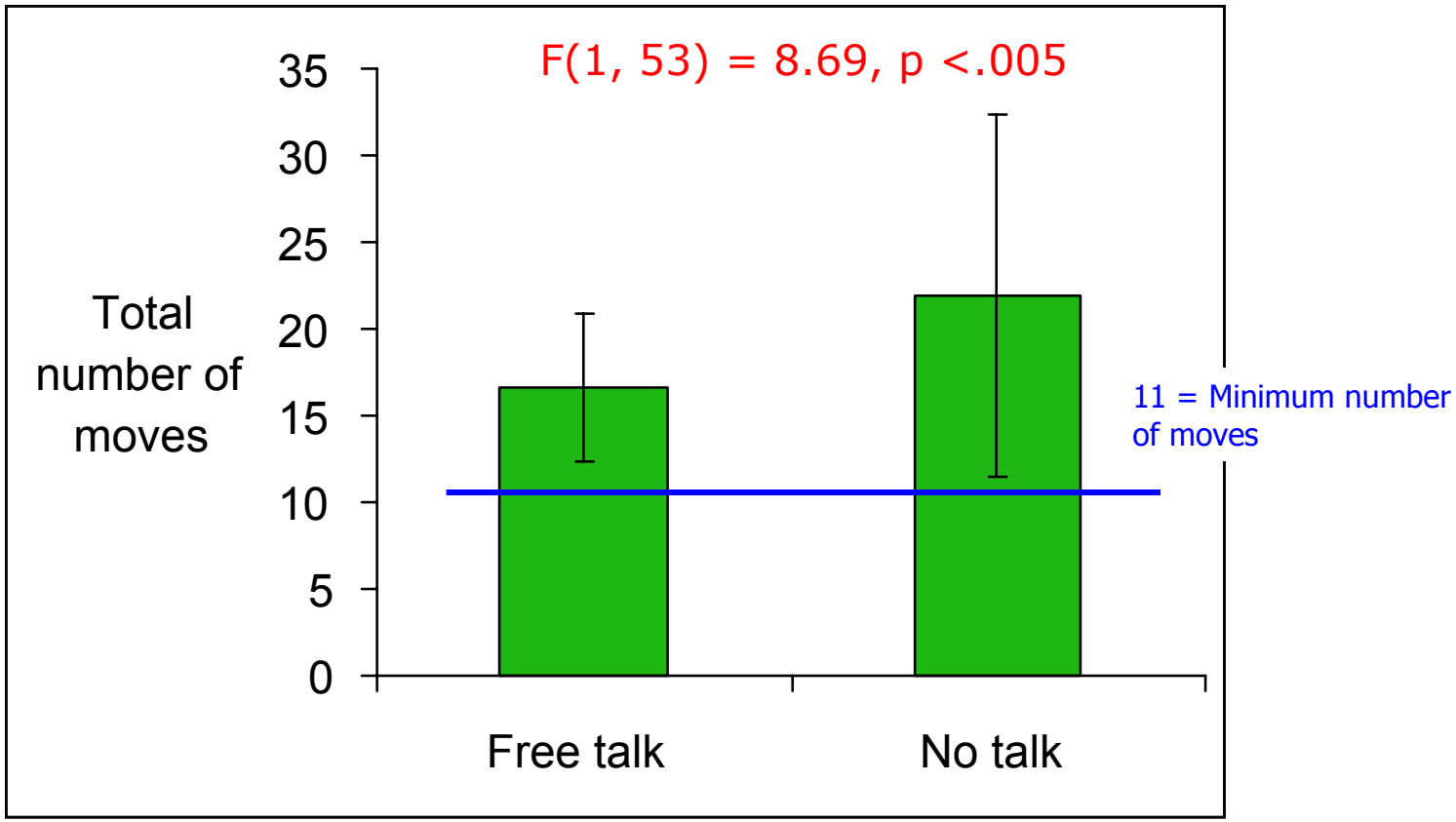
Main effect of trust on time on task

- Teams that trusted needed less time to complete the task:



Main effect of communication mode on total number of moves

- Teams that communicated freely (whether face-to-face or by radio) required fewer moves to complete the task:



Revealed trust ➔ Team performance

- Both trust and communication mode had a significant impact on team performance in this cooperative problem-solving task.
  - Trust and communication did not interact.

- Trust improves team effectiveness (less time)
  - vs. distrust and betrayal.
- Communication improves team efficiency (fewer moves).

- Proximity had no effect.
  - Radio communication does not degrade cooperative problem solving.

Revealed trust ⊘ Individual performance

- The category of revealed trust did not influence individual cognitive performance in the mental rotation task.
- Players were both faster and more accurate in the second trial, $p < .001$.

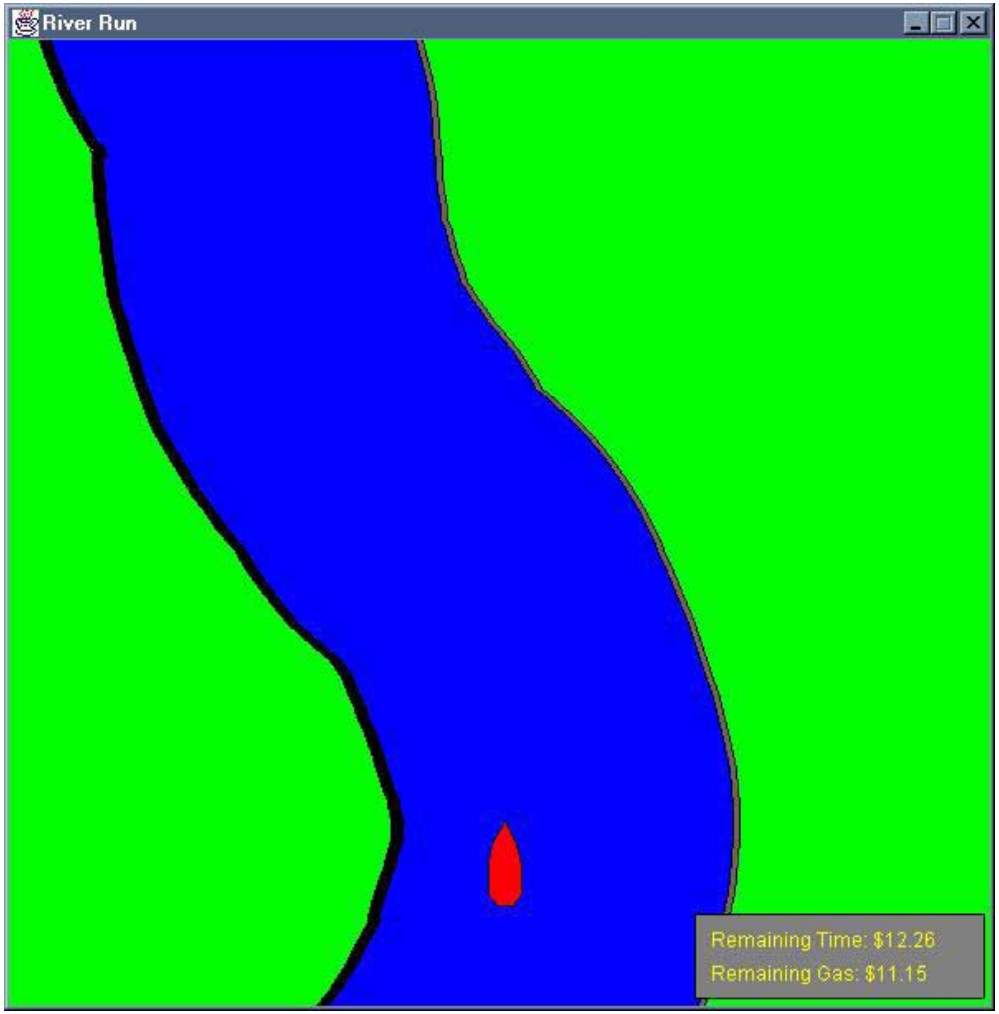
Experiment 2: F game ➔ Boat game

On the impact of revealed trust on performance in a distributed command and control task

The boat game

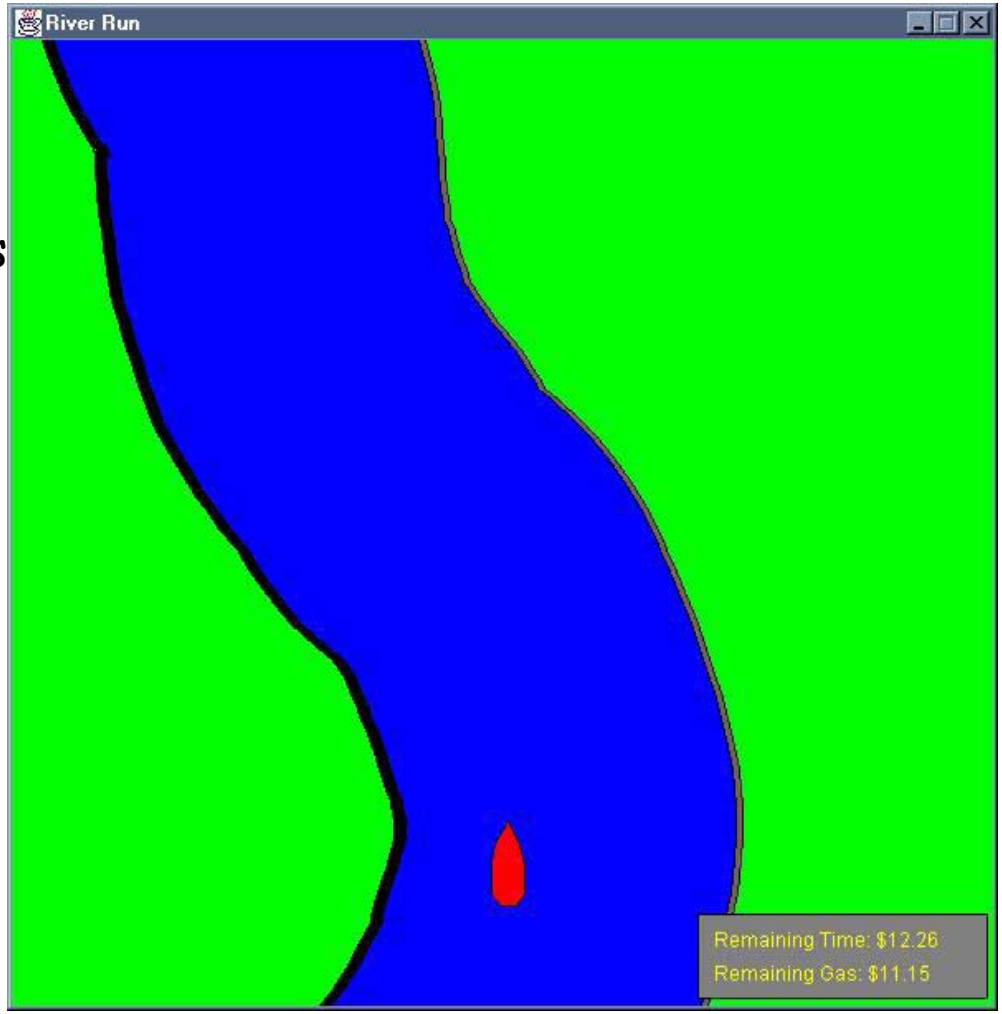
- Task: Work together to navigate a tug boat up a sinuous river to the dock.
- Player 1 is the pilot, responsible for steering.
- Player 2 is the engineer, responsible for speed.

- Running into the bank is not good.
- The boat has inertia and is slow to respond to commands.



The boat game

- Dependent measures:
  - Time on task
  - Number of errors (crashes into the bank).
- Incentives:
  - Split $ total shown on two meters when (if) the boat reaches the dock.



Experiment 2: The boat game

- Question:
  - How do revealed trust, proximity, and communication mode interact to affect team cognitive performance in a distributed command and control task?

- Hypotheses:
  - We expect cognitive performance to improve if:
  - A) Participants are permitted to communicate freely,
  - OR
  - B) They work face-to-face rather than over a radio link,
  - OR
  - C) They are both trusting and trustworthy.

Participants, design, procedure

- 63 pairs of participants (19 female pairs, 17 male pairs, and 27 mixed pairs; median age 19, mode 18, maximum 28).

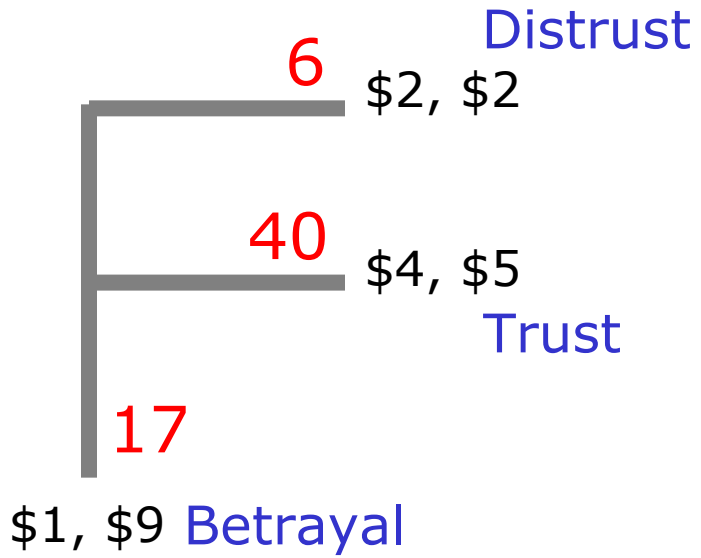
- 2 X 2 X 3 between-subjects (teams) design
  - Proximity X Communication X Revealed trust

  - Again, trust was a quasi- (independent) variable at three levels. Pairs in all 4 levels of proximity and communication mode played the F game to self-select into one of three categories of trust: distrusting, trusting/trustworthy, betraying prior to playing the boat game.

Results: F game

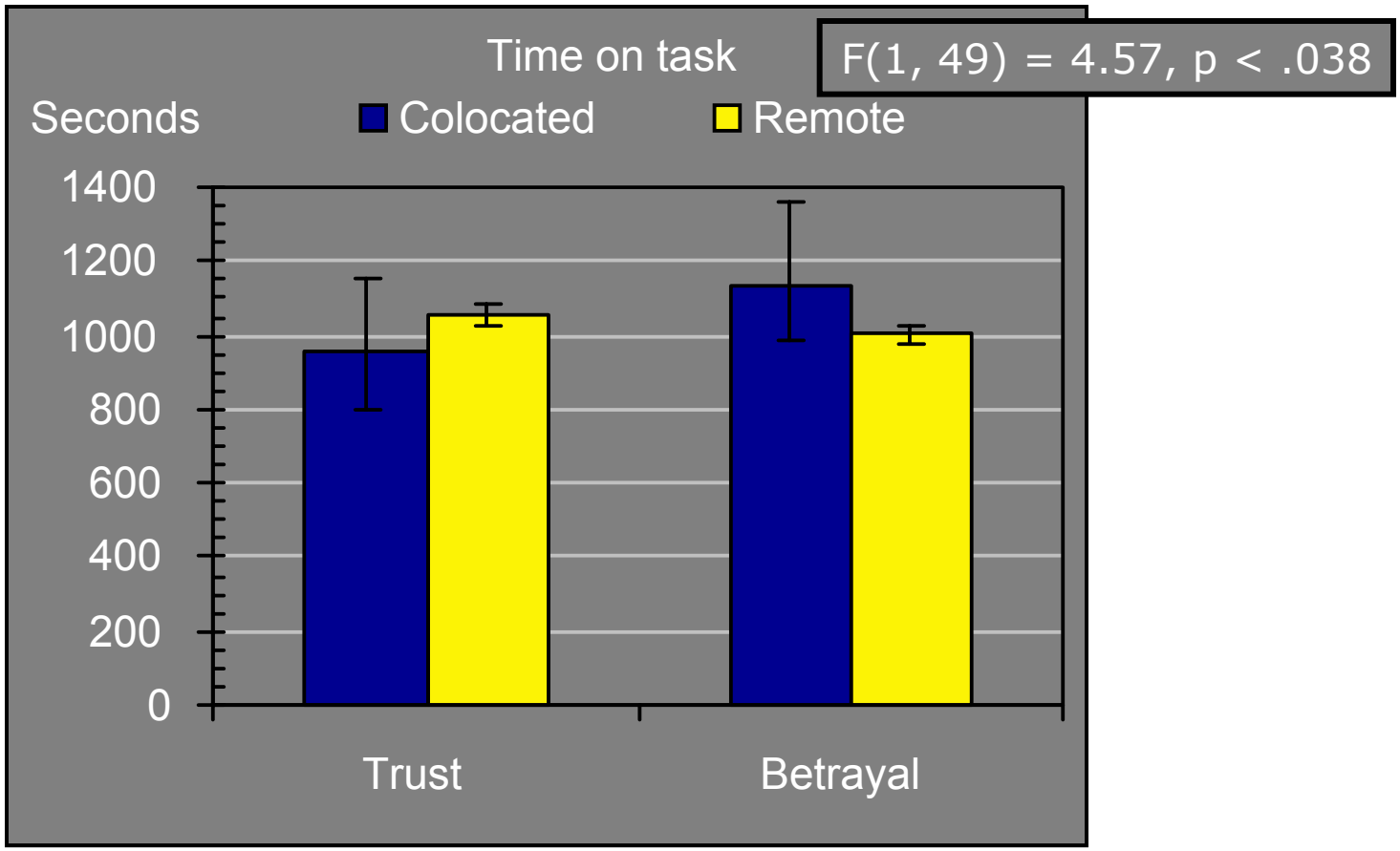
- N outcomes
- Only 6 teams experienced distrust, an insufficient number for statistical power. They have been excluded from the analysis.

- We report the results of the 2x2x2 ANOVA on the remaining 57 teams where trusting and betraying are the two trust conditions.



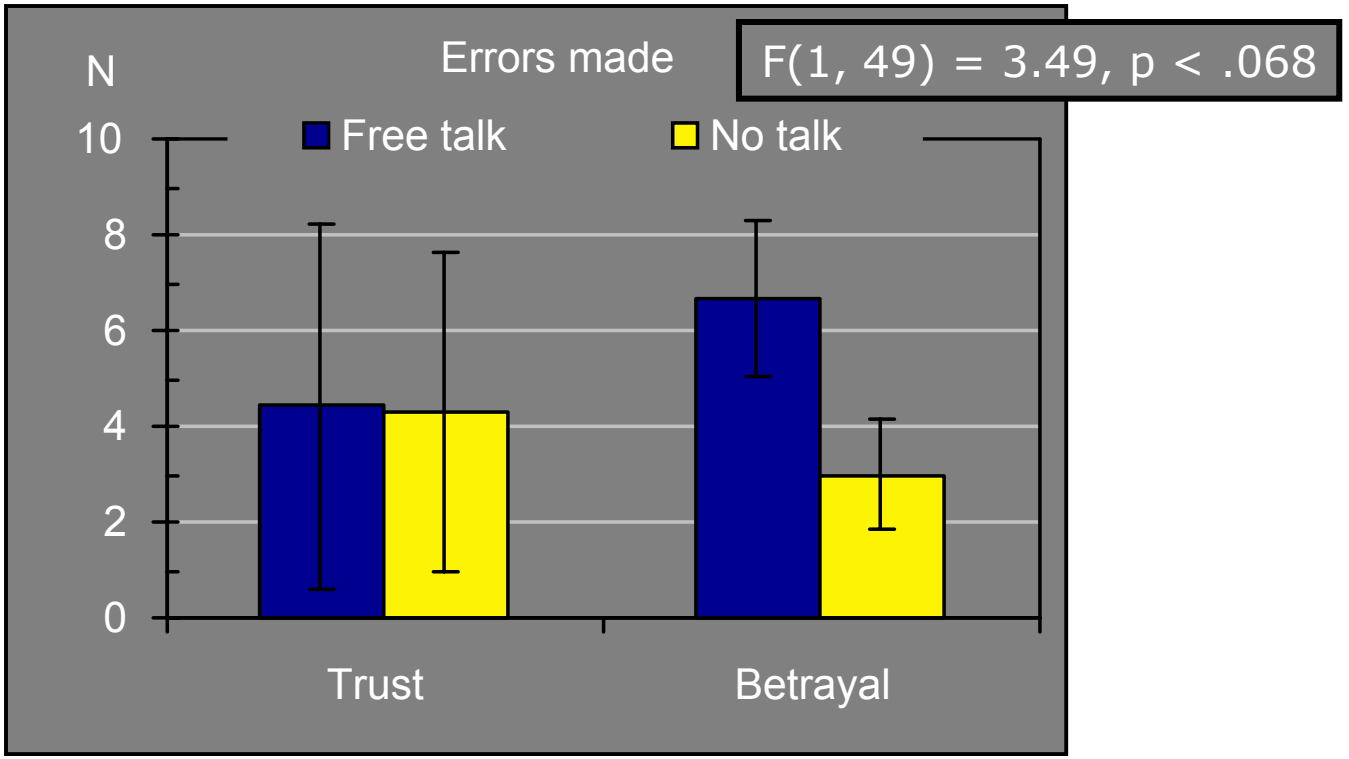
Trust X Proximity interaction

- Betrayal hampers efficiency when teams work face-to-face but not when they work remotely:



Trust X Communication mode interaction

- Communication mode has no effect on error rate for teams that trusted and were trustworthy.
- Teams that experienced betrayal were less effective if, and only if, they had to communicate:



Revealed trust ➔ Team performance

- Revealed trust interacts with both proximity and communication mode to influence performance in this distributed command and control task.

- The failure to reciprocate an expression of trust with trustworthiness can have a profoundly negative influence on (1) how well teams communicate and on (2) how well they coordinate their activities when working face-to-face.

Revealed trust ⊘ Individual performance

- Once again, the category of revealed trust did not influence individual cognitive performance in the mental rotation task.
- Players were both faster and more accurate in the second trial, $p < .001$.

Experiment 3: Boat game ➔ F game

On the impact of performance in a distributed command and control task on revealed trust

Method, Experiment 3

- In experiment 3, the team task precedes the F game:
  - Individual cognitive task
  - Team task (boat game)
  - Trust (using the F game) is a dependent variable
  - Repeated play of individual task

Boat game ➔ F game

- Question:
  - How do proximity, communication mode, and cognitive performance in a distributed command and control task interact to affect subsequent trusting behavior?

  - This experiment turns the question around to assess the direction of causation between trust and performance.

Boat game ➔ F game

- Hypotheses:
  - We expect the likelihood of trust to increase if:
  - A) Participants are permitted to communicate freely,
  - OR
  - B) They work face-to-face rather than over a radio link,
  - OR
  - C) They complete the task successfully.

Design, procedure, participants

- 2 X 2 between sjs (teams) design
  - Proximity X Communication

- Trust is a dependent variable at three levels.
  - After finishing the boat game, pairs played the F game to self-select into one of three categories of trust: distrust, trust/trustworthy, betrayal.
  - Analyses require logistic regression.

- 63 pairs of participants (19 female pairs, 17 male pairs, and 27 mixed pairs; median age 19, mode 18, maximum 40).

Null results for performance

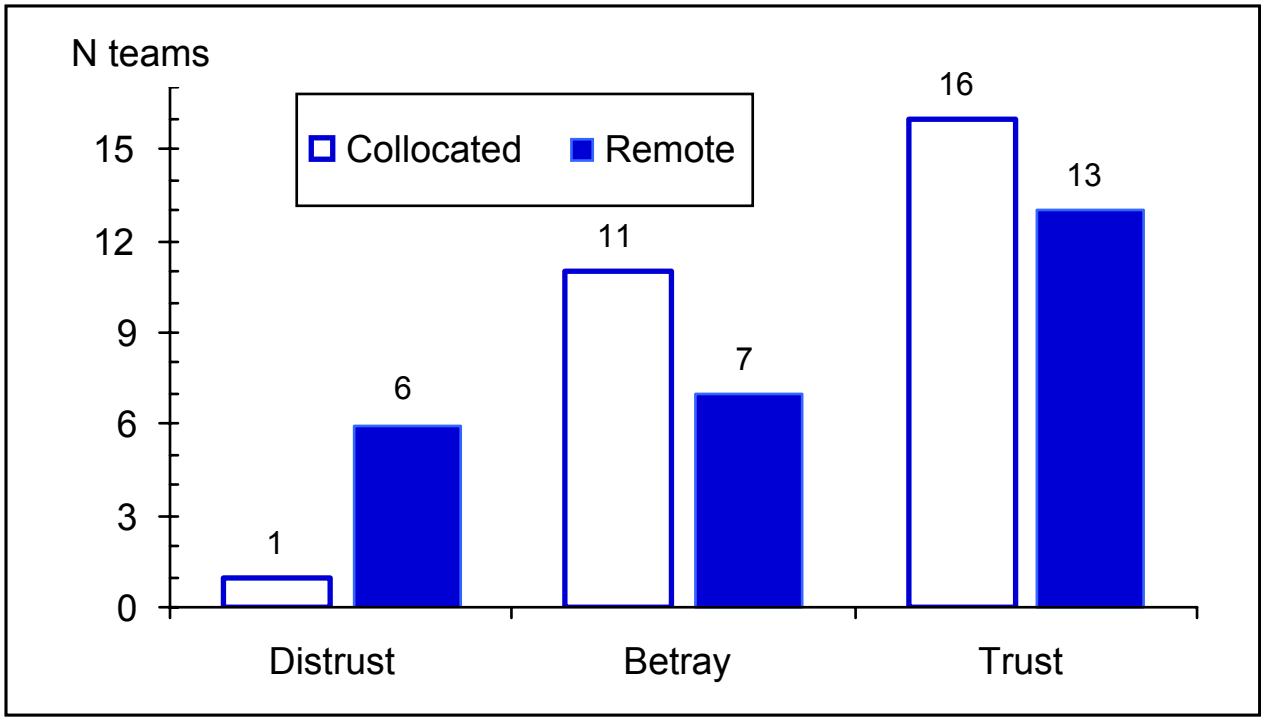
- As in Experiment 2, communication mode and proximity had no effect on team performance (errors made and time on task).

- Quality of team performance had no effect on revealed trust.

Team performance  Revealed trust

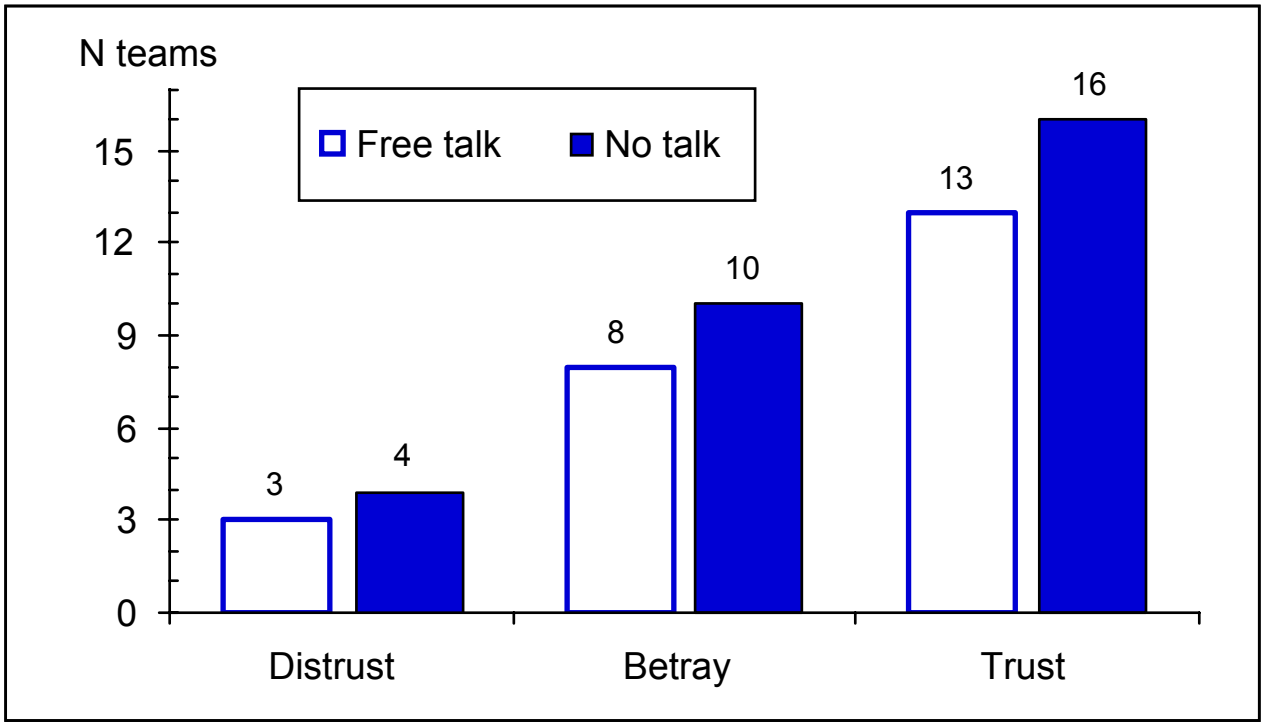
Main effect of proximity on revealed trust

- Proximity has a strong effect on revealed trust:
  - Teams that worked face-to-face were more likely to trust and less likely to distrust.



$X^2(2) = 7.85, p < .02$

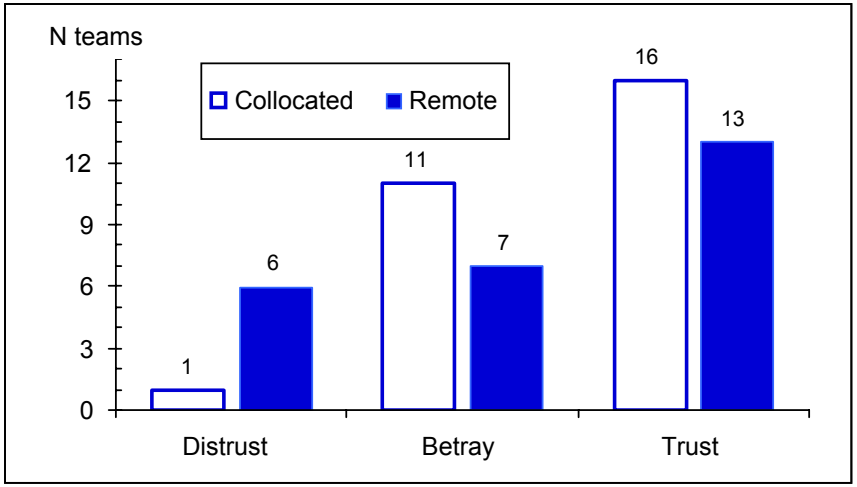
No main effect of communication on revealed trust

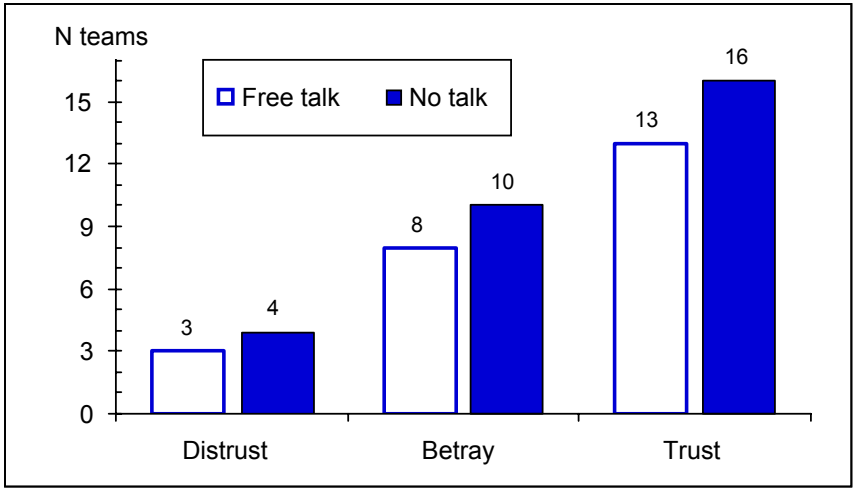


$X^2(2) = 0.22, p > .89$

Logistic regression: Betrayal vs. Trust

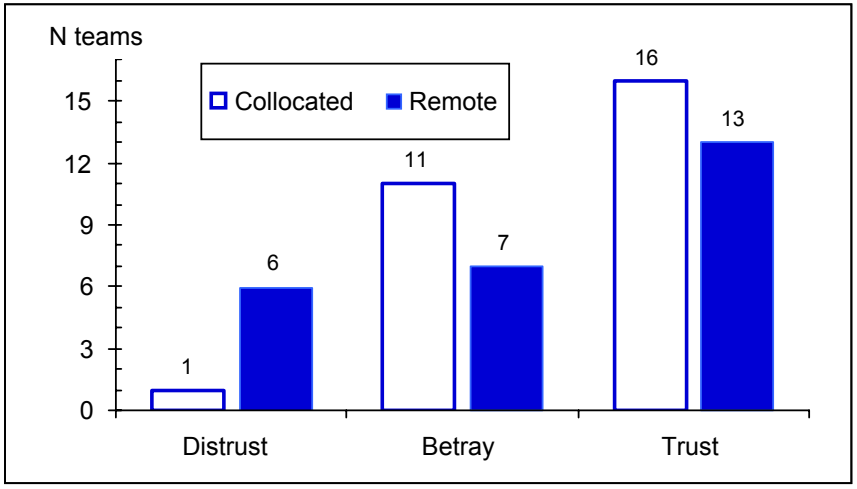
- L(B/T) = -0.68 - 0.22 Remote - 0.16 Free talk
  - The likelihood of Betrayal decreases when in the remote and the free talk conditions

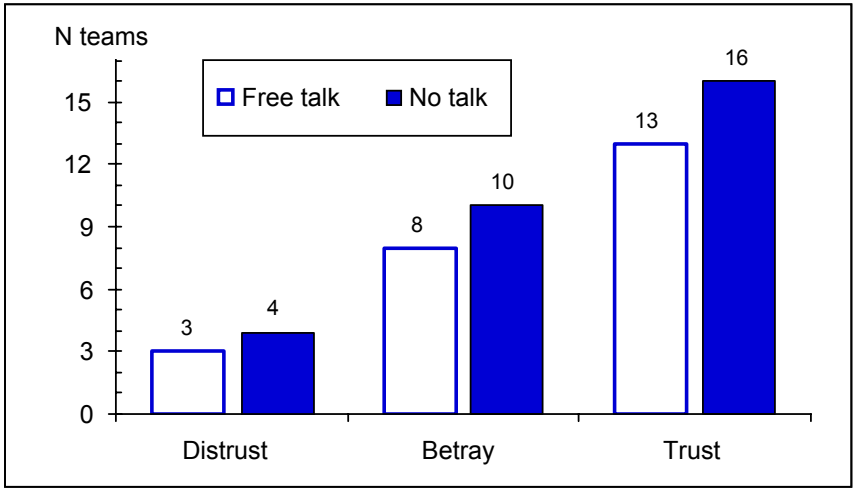




Logistic regression: Distrust vs. Trust

- L(D/T) = -5.90 + 5.09 Remote (- 0.08 Free talk)
  - The likelihood of Distrust increases when in the remote (and decreases marginally in the free talk) conditions





Proximity ➔ Revealed trust

- Proximity played a critical role in the formation of trust between members of our teams. Teams that were colocated as they worked together were more likely to trust and be trustworthy.

- This does not bode well for organizations that form teams with members in dispersed locations.

- When a team is initially assembled, it would be wise to bring all members physically together to facilitate the common ground that underlies trust and to offset the potential for betrayal.

Summary

- Exp. 1: Revealed trust ➔ Cooperative problem solving
  - Trust improves effectiveness (less time).
  - Communication improves efficiency (fewer errors).
  - No effect of proximity
- Exp. 2: Revealed trust ➔ Distributed C^2
  - Betrayal x proximity interaction
  - Betrayal x communication interaction
- Exp. 3: Proximity ➔ Revealed trust
  - Only proximity played a critical role in the formation of trust.

Returning to the issues

- What is the role of trust in distributed command and control?
  - Does trust influence team performance ?
  - Trust improves effectiveness but not necessarily efficiency.
  - Performance may be significantly degraded when trust is not reciprocated with trustworthiness.

  - Does team performance influence trust ?
  - Not that we can see.

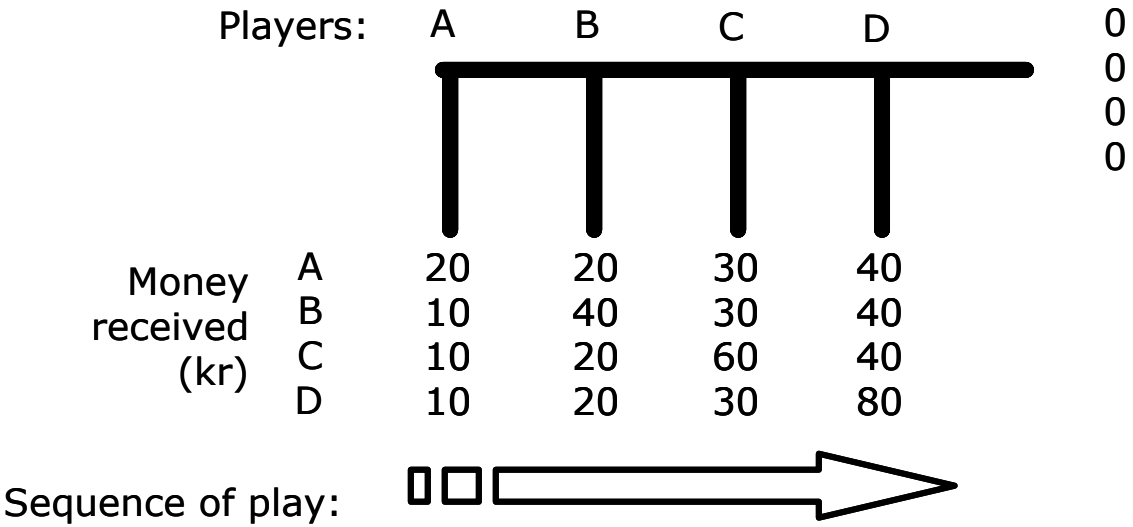
  - What are the effects / interactions of trust with proximity and communication?
  - Proximity facilitates the formation of trust.

Current and planned laboratory experiments

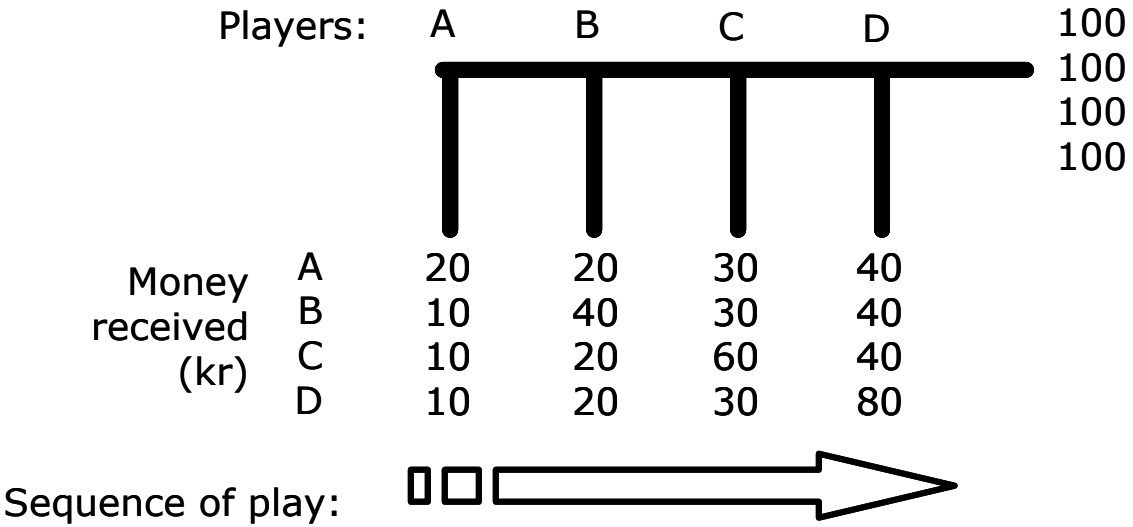
- Replicating all three experiments with Swedish participants.
  - Cultural differences in revealed trust?

- Test hypotheses about the utility of the centipede game as a tool for establishing common ground and instilling trust.

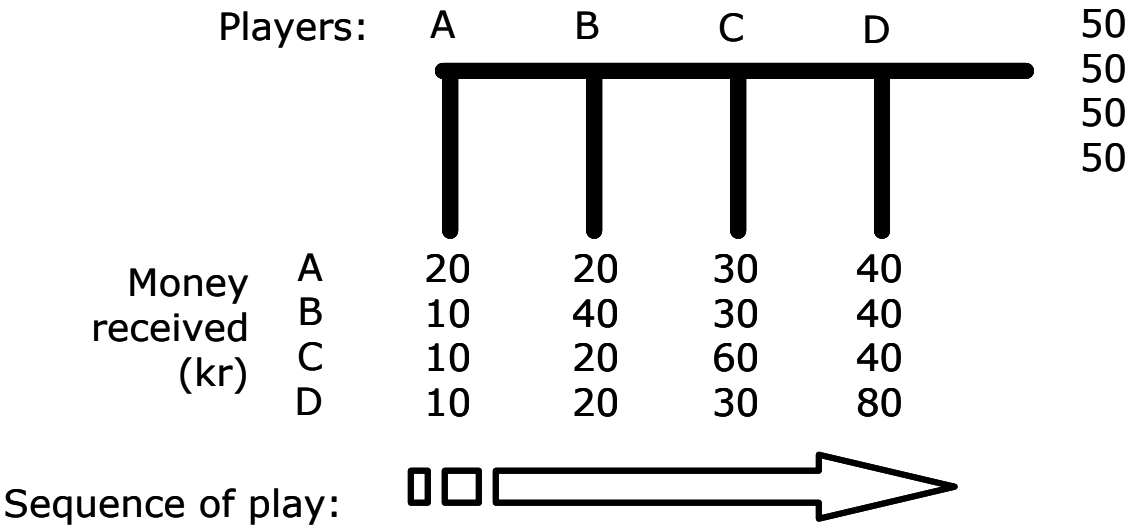
The traditional centipede game



The mutual cooperation game



A tool for instilling trust?



Student Collaborators

- Shala Bloomberg
- Jessica Bolton
- Nathan Fosse
- Steve Freund
- Whitney Gee
- Jill Halleran
- Serena Hawkins
- Andy Kearney
- Holly Pangburn
- 2Lt Keith Pangburn
- Nicole Pruitt
- Liz Ratcliff
- Andrea Shane
- Jana Smith
- Julie Soupene
- Ben Swan
- Molly Toll
- Beth Younger

Acknowledgements

- This work was supported by the DOD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621.
- Dr. Elmar Schmeisser is the project monitor. Prof. Peter Hancock is the project principal investigator.
- The views expressed in this work are those of the author and do not necessarily reflect official Army policy.



Linköping University

Progress Report covering the period 1 July 2003 – 30 September 2003
MURI OPUS -- DAAD19-01-1-0621

Principal Investigator: Kip Smith, Ph.D.
Division of Industrial Ergonomics

This is the first progress report to be filed from Linköping University (LiU) as part of the MURI OPUS research program. The second quarter progress report (submitted from Kansas State University, KSU) enumerated six goals for the second half of 2003. The first three of the six goals were met during the third quarter:

√ Recreate and expand the KSU laboratory experiments at LiU.

√ Add psychophysiologic data collection for assessments of stress.

√ Establish research links with the Swedish military and other institutions.

- Refine experiments designed to test hypotheses about the centipede game as a mitigator of the stress associated with the experience of distrust.
- Recreate and expand the KSU paintball experiments at LiU.
- Extend the paintball paradigm from trust-by-individuals-in-a-leader to trust-by-multi-person-teams-in-a-leader and to trust-among-team-members.

Specifically, the principal investigator (PI) and all MURI OPUS-supported hardware and software moved successfully, without breakage or loss, from KSU to LiU. At LiU, permissions were obtained to conduct MURI OPUS-supported experimental research. The laboratory experiments have been successfully restarted at LiU. In addition, an opportunity for collaboration has been forged at the highest level of the Försvarshögskolan, the Swedish National Defense College. This memo discusses these accomplishments in turn and establishes target goals for the next six months.

Transition to LiU

The PI spent most of the quarter moving to Sweden, settling in at LiU, obtaining a Swedish 'person-number' (without which one is merely a tourist and cannot draw a salary or receive extramural funding), and obtaining the requisite permissions to conduct experimental research with human subjects. There were many barriers to this process which had to be crossed in the correct sequence.

As a result, the PI is now legally able to (1) earn a living in Sweden and to (2) conduct experimental research with human subjects using the paradigms discussed in the second quarter 2003 progress report.

The psychophysiologic data recording equipment and associated software made it through Swedish customs with only minor hitches. The cost of import have been absorbed by the PI as part of cost of moving his family and personal belonging to Sweden. The equipment is up and running in the PI's laboratory at LiU. Data collection is underway. Thus, the second of the six goals listed above has been met.

The only part of the project that has yet to make it to LiU is the funding. As of 30 September, 2003, LiU has not received a finalize subaward agreement from the University of Central Florida (UCF). Much of the delay is due to spreadsheet errors on the proposed budget pages that were made by the PI and that were not caught and corrected before they had been passed on to the DOD. The PI assumes full responsibility for this unfortunate turn of events.

Laboratory quasi-experiments

Permission to conduct laboratory experiments on the effects of trust and distrust on team cognitive performance has been obtained from the LiU Ethical Research Committee and forwarded to UCF.

The quality of and access to computer facilities at LiU vastly outstrips what was available at KSU. The PI's laboratory at LiU contains 17 computers that are configured in a server-client network to support the data acquisition from as many as eight dyads (teams of two) simultaneously. Data acquisition has begun at LiU using the multi-task experimental paradigms described in the second quarter 2003 progress report. Thus, the first of the six goals listed above has been met.

The experiments currently in progress at LiU aim to test their generality of our previously (published) findings of a differential impact of trust and distrust on team cognitive performance. In addition, they are confirming, cross-culturally, the utility of the economic trust game as a quasi-manipulation of trust and distrust.

Showing that Swedish participants respond like American participants to the quasi-manipulation of trust and distrust is the first step in a proposed series

of experiments. If the cross-cultural differences are negligible or tractable, then we will be able to proceed with the series of experiments identified in the fourth goal.

These experiments will (1) use the extensive form trust game to get participants to reveal their dispositions to trust or distrust, (2) document the team cognitive decrements associated with distrust, and (3) test hypotheses about the utility of a proposed mitigator of the stress associated with the experience of distrust. The mitigator is a second economic game known as the centipede game. This game and the full experimental design are discussed in the LiU proposal for funding which was sent to UCF in June, 2003.

A graduate student has been found who is willing and able to write the Java code for the centipede game. The PI is unable to employ a student until LiU and UCF sign the subaward contract. Once the contract is signed, progress towards this goal will be relatively quick.

Contacts with the Swedish National Defense College

The PI has had several conversations with Professor Berndt Brehmer, Professor of Command and Control Science at the Försvarshögskolan (the Swedish National Defense College) in Stockholm. Professor Brehmer is an internationally renown scholar and an expert in dynamic and distributed decision making. He is keen on collaborating with the PI on issues such as the impact of trust and distrust on team performance in distributed warfare centers. Such a collaboration would complement the PI's MURI OPUS-supported work on the impact of trust and distrust on dyad performance and on the willingness of soldiers to follow commands issued by remote leaders. Thus, progress is being made on the third of the six goals listed above.

Field experiments (paintball)

During the third quarter, the PI had several conversations with the operators of commercial paintball facilities in Linköping about the feasibility and costs of setting up an assault lane and conducting experiments in their facility. These conversations have led the PI to expect to be able to conduct the field experiments indoors during the winter of 2003-2004. These experiments will investigate the impact of remote command and control on soldiers' willingness

to obey commands while under live paintball fire. It is anticipated that the first set of experiments will be much like those conducted at Fort Riley, Kansas.

Further progress toward this goal has been hampered by the fact that paintball facilities in Sweden are operated seasonally, during the dark Swedish winter. The only impediments to data acquisition appear to be the lack of cash (no UCF – LiU subaward yet, see above) and the time of year. Nevertheless, incremental progress is being made on the fifth of the six goals listed above.

Permission to conduct experiments with live paintball fire has been obtained from the LiU Ethical Research Committee and forwarded to UCF.

Goals for the fourth quarter 2003 and the first quarter 2004

The long-term goals of the LiU portion of the MURI OPUS project is to identify and mitigate team cognitive decrements due to distrust and to remote command and control.

The goals for next six months are to:

- Solidify research links with the Försvarshögskolan and other institutions.
- Replicate and extend the laboratory experiments using the extensive form trust game.
- Implement several versions of the centipede game in Java code.
- Conduct preliminary tests of the utility of the centipede game as a mitigator of the stress associated with the experience of distrust.
- Organize, replicate, and extend the field experiments at a local paintball facility.
- Extend the paintball paradigm from trust-by-individuals-in-a-leader to trust-by-multi-person-teams-in-a-leader and to trust-among-team-members.

CEREBRAL VASCULARITY AND PERFORMANCE ON AN ABBREVIATED VIGILANCE TASK

Todd D. Hollander[1], William S. Helton[2], Lloyd D. Tripp[1], Kelley Parsons[1], Joel S. Warm[1], Gerald Matthews[1], William N. Dember[1], and Raja Parasuraman[3]

[1]University of Cincinnati
Cincinnati, Ohio

[2]Wilmington College
Wilmington, Ohio

[3]Catholic University of America
Washington D.C.

Transcranial Doppler sonography (TCD) and transcranial cerebral oximetry (TCCO) measures of blood flow and oxygenization levels in the brain were collected while observers performed an abbreviated 12 min vigilance task designed by Temple et al. (2000) to serve as an analog to more traditional long-duration vigils. Both measures showed higher levels of cerebral vascular activity in the right as compared to the left cerebral hemisphere indicating that the overall level of performance in the abbreviated vigil is right-lateralized, a finding that coincides with the outcome of earlier blood flow studies featuring more traditional long-duration vigils. This parallel provides strong support for Temple et al's (2000) argument that the abbreviated vigil is a valid analog of more traditional long-duration vigilance tasks and implies that laterality in vigilance is a generalized effect in terms of cerebral vascular dynamics that appears in terms of both blood oxygenation and hemovelocity.

Several investigations using positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) procedures have been successful in demonstrating that changes in cerebral blood flow and glucose metabolism are involved in the performance of sustained attention or vigilance tasks (Parasuraman, Warm, & See, 1998). These studies identified multiple brain regions that are active during the performance of such tasks, including the nucleus locus coerulus, the brainstem reticular formation, the midbrain tegmentum, the intralaminar region of the thalamus, the cingulate gyrus, and the frontal lobe. However, as Parasuraman et al. (1998) emphasized, an important limitation of these studies is their failure to correlate brain activity with performance efficiency, perhaps due to the high costs and restrictive environments associated with PET and fMRI use.

The high costs and the restrictions of PET and fMRI may be circumvented in studying brain systems in vigilance by employing transcranial Doppler sonography (TCD) – a relatively inexpensive and noninvasive procedure that allows for continuous monitoring of cerebral blood flow (CBF) in the main-stem intracranial arteries. When an area of the brain becomes metabolically active, as in the performance of mental tasks, by-products of this activity, such as CO_2, increase. This results in elevation of blood flow to the region to remove the waste product (Aaslid, 1986). Thus, CBF may be viewed as a metabolic index of information processing (Stroobant & Vingerhoets, 2000). Along this line, a recent series of studies (Hitchcock, 2003; Hollander et al., 2002; Mayleben, 1998) has indicated that the vigilance decrement, the decline in signal detections over time that typifies vigilance performance, is paralleled by a decline in CBF. In addition, the absolute level of blood flow in these studies was closely related to the psychophysical and cognitive demands of the vigilance task, the availability of reliable cues to the imminent appearance of signals to be detected, and to the ability of observers to form accurate expectancies as to the time of signal arrival. All of these overall effects were lateralized to the right cerebral hemisphere, consistent with PET and FMRI studies indicating right hemisphere functional control of vigilance performance (Parasuraman et al., 1998).

In addition to the TCD procedure, there is another nonrestrictive and relatively economic alternative to PET and fMRI -- the measurement of cerebral blood oxygen saturation using near infrared spectroscopy or transcranial cerebral oximetry (TCCO; Toronov et al., 2001). This procedure provides information pertaining to regional oxygenation of cerebral tissue, and results with

this technique show that tissue oxygenation increases with the information processing demands of the task being performed by an observer (Punwani et al., 1998; Toronov et al., 2001). Hence, one might expect that along with CBF, cerebral oxygenation would also be related to vigilance performance. An examination of that possibility was one goal for the present study.

Because of their demanding character, vigilance tasks have been used in test batteries to assess the effects of stressors, such as infectious disease and drugs, on performance (Alluisi, 1969; Damos & Parker, 1994) and to examine the impact of a variety of clinical entities on attention (Berch & Kanter, 1984). Although vigilance tasks have been useful in these roles, their long duration has often been an inconvenience. Posner (1978) has suggested that task duration in vigilance is an arbitrary dimension and that it should be possible to develop shorter-duration tasks that exhibit the same effects as the typical long-duration vigils. Toward this end, Temple et al. (2000) developed an abbreviated (12 min) vigilance task that shows the vigilance decrement and mirrors long duration tasks in regard to psychophysical and psychopharmacological variables. As in the case of long-duration vigils, performance in the abbreviated vigil varies inversely with signal salience and is enhanced by caffeine (Temple et al., 2000). The demonstration that the cerebral vascular dynamics associated with the abbreviated vigil duplicate those associated with long-duration vigils would provide additional evidence that the former represents a viable analog to the latter. Toward that end, the present study employed both the TCD and TCCO procedures with the abbreviated vigil.

Experiment 1: The TCD measure

Methods

Six observers (3 males and 3 females) were assigned at random to each of four conditions resulting from the factorial combination of vigilance condition (work-imperative and control) and hemisphere (right and left). In the vigilance condition, observers inspected the repetitive presentation on a VDT of light gray capital letters consisting of an "O," a "D," and a "backwards D." The letters were exposed for 40 msec at a rate of 57.5 events/min against a visual mask consisting of unfilled circles on a white background. Critical signals for detection (p = 0.20/period of watch) were the appearance of the letter "O." Observers signified their detection of critical signals by pressing the key on a response pad. Prior to the main vigil, participants in the vigilance condition were given a 2-min. period of practice to familiarize themselves with the task. Participants in the control condition were required to simply stare at the computer screen for 12-min. with no work imperative. Blood flow determined hemovelocities in the left and right middle cerebral arteries (MCA's) were measured during a 5- minute baseline rest period and throughout the sustained attention task using a Nicolet/Multi-Dop X4 TCD unit. The mainstem MCA, which was monitored at depths of 45-55mm (distance between the transducer face and sample volume), typically has a hemovelocity range of 50-65 cm/sec. The TCD unit permitted depth adjustment in 5mm increments as needed for isolating the MCA. As is typical of research using the TCD technique (Aaslid, 1986), hemovelocity during the final 60 seconds of the baseline period provided the baseline index.

Results

Correct Detections. Mean percentages of correct detections in the vigilance task is plotted as a function of periods of watch in Figure 1.

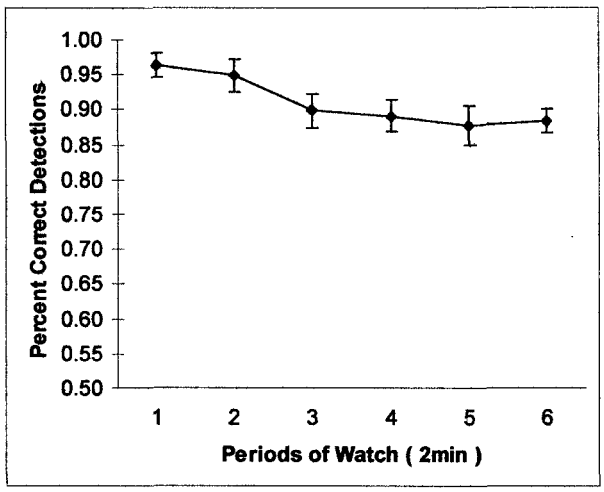


Figure 1.

A repeated measures analysis of variance (ANOVA) of an arcsin transformation of the detection scores revealed that there was a significant decline in signal detections over the total watch F (4, 39) = 3.91, p = .012.

Cerebral Blood Flow. A 2 (condition) x 2 (hemisphere) x 6 (periods of watch) split-plot analysis of variance (ANOVA) based on the percentages of hemovelocity change from baseline revealed that significantly higher blood flow levels occurred in the right hemisphere (M = 1.01) than in the left hemisphere (M = 0.98), F (1, 20) = 5.06, p = .036, and a significant condition x hemisphere interaction, F (1, 20) = 4.65, p = .043. All other effects lacked statistical significance (p > .05).

Mean percentages of hemovelocity change from baseline for the task and control conditions are plotted for both hemispheres in Figure 2.

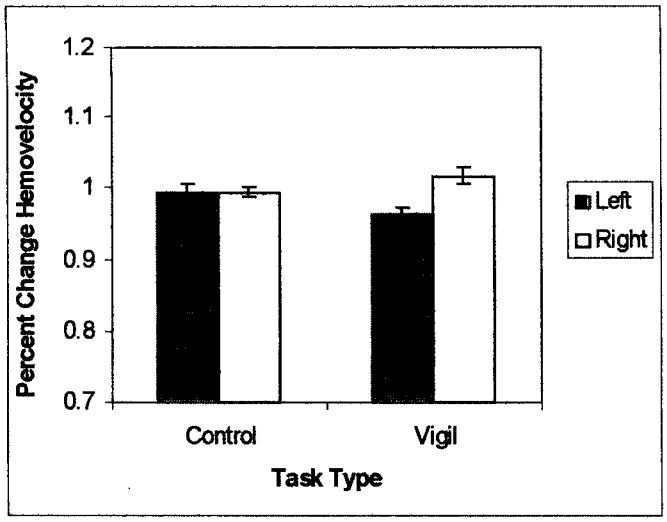


Figure 2.

In order to further investigate the Condition x Hemisphere interaction, paired t-tests between the mean levels of percent hemovelocity change in the right and left hemispheres were conducted for both the vigilance task and control groups. There was a significant difference between the right (M = 1.02) and the left (M = 0.96) hemisphere for the vigilance task group, t (10) = 3.28, p = .008. For the control group, however, the mean difference between the right (M = 0.99) and the left (M = 0.99) hemisphere was not significant, t (10) = .062, p = .952.

Experiment 2: TCCO Measures

Method

Nineteen right-handed observers were assigned at random to one of two conditions: a vigilance condition and a control condition of equal duration in which there was no work imperative. As in the first experiment, control observers were required to simply stare at a computer screen for 12-min. TCCO measures from the frontal lobe were provided by a Somanetics (Troy, MI) INVOS 4100 Cerebral Oximeter. The frontal lobe was employed because of PET and fMRI studies showing right dominance in this brain region during the performance of vigilance tasks (Parasuraman et al., 1998). Cerebral oxygen saturation data were collected for three minutes prior to the experimental session in order to provide a baseline comparison.

Results

Correct Detections. Mean percentages of correct detections for the vigilance task are plotted as a function of periods of watch in Figure 3.

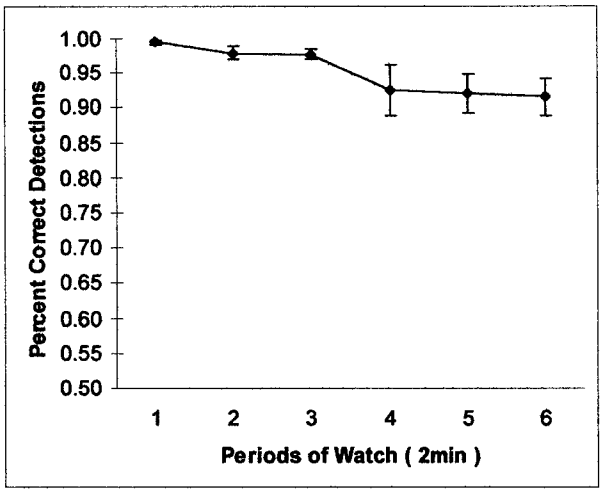


Figure 3.

A repeated measures analysis of variance (ANOVA) based on an arcsin transformation of the detection scores revealed that there was a significant decline in signal detections over the total watch F (5, 90) = 4.92, p = .001.

Cerebral Oxygen Saturation. A 2 (condition) x 2 (hemisphere) x 6 (periods of watch) split-plot analysis of variance (ANOVA) based on the percentages of oxygen change from baseline revealed that significantly higher oxygen saturation levels occurred in the vigilance task condition (M = 4.02) than in the control condition (M = 2.02), F (1, 36) = 5.74, p = .02, and that there were significantly higher oxygen saturation levels in the right hemisphere (M = 3.45) than in the left (M = 2.60), F (1, 36) = 4.84, p = .03. There was, moreover, a significant condition x hemisphere interaction, F (1, 36) = 4.01, p = .05. All other effects lacked statistical significance (p > .05).

Mean percentages of oxygen saturation change from baseline for the task and control conditions are plotted for both hemispheres in Figure 4.

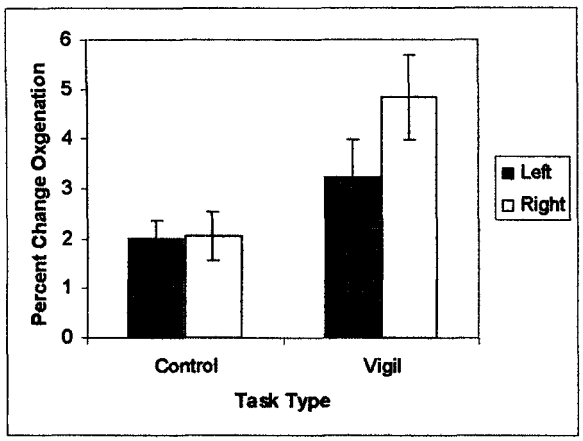


Figure 4.

In order to further investigate the Condition x Hemisphere interaction paired t-tests between the mean levels of percent oxygen change in the right and left hemispheres were conducted for both the vigilance task and control groups. There was a significant difference between the right ($M = 4.84$) and the left ($M = 3.21$) hemisphere for the vigilance task group, $t = 2.29$, $p = .03$. For the control group, however, the mean difference between the right ($M = 2.06$) and the left ($M = 1.98$) hemisphere was not significant, $t = .246$, $p = .81$.

Discussion

In this investigation, blood flow velocity and oxygenation were both found to be significantly higher in the right than in the left cerebral hemisphere among observers who performed the vigilance task. In contrast, hemispheric differences with either measure were not found among control observers who had no work imperative. Clearly, the results of both experiments indicate that the overall level of performance in the abbreviated vigil is right lateralized, a finding that coincides with the outcome of earlier blood flow studies featuring more traditional long-duration vigils (Hitchcock et al. 2003; Hollander et al., 2003; Mayleben, 1998) and with PET and fMRI investigations (Parasuraman et al., 1998). This parallel has several important implications. It provides strong support for Temple et al's (2000) argument that the abbreviated vigil is a valid analog of more traditional long-duration vigilance tasks. The fact that the TCCO procedure yielded similar laterality effects as the TCD procedure further implies that laterality in vigilance is a generalized effect that appears in terms of both blood oxygenation and hemovelocity. It also implies that the TCCO procedure may be a useful supplement to the TCD approach in providing a noninvasive imaging measure of brain activity in the performance of a vigilance task.

In contrast to the laterality effect in the overall level of performance, the results with regard to the vigilance decrement did not coincide with earlier findings from long-duration tasks that the decrement is accompanied by a decline in cerebral blood flow (Hitchcock, 2003; Hollander et al., 2002; Mayleben, 1998). Although there was a significant decline in signal detections over time in both experiments, neither the blood flow nor the blood oxygenation measure changed significantly with time on task. It is possible that cerebral vascular dynamics are structured so that overall hemispheric dominance emerges early in the time course of task performance but that temporally based declines in cerebral vascular activity require a considerable amount of time to become observable. Thus, the abbreviated 12-min vigil employed in this study, which was only about 30% as long as the earlier blood flow studies, may not have been long enough to permit time-based declines in cerebral vascular dynamics to be observed.

References

Aaslid, R. (1986). Transcranial Doppler examination techniques. In R. Aaslid, (Ed.), *Transcranial Doppler sonography* (pp. 39-59). New York: Springer-Verlag.

Alluisi, E.A. (1969). Sustained performance. In E.A. Bilodeau & I.M. Bilodeau (Eds.), *Principles of skill acquisition* (pp. 59-101). New York: Academic Press.

Berch, D.B., & Kanter, D.R. (1984). Individual differences. In J.S. Warm (Ed.), *Sustained attention in human performance* (pp. 143-178). Chichester, UK: Wiley.

Damos, D.L., & Parker, E.S. (1994). High false alarm rates on a vigilance task may indicate recreational drug use. *Journal of Clinical and Experimental Neuropsychology*, 16, 713-722.

Hitchcock, E.M., Warm, J.S., Mathews, G., Dember, W.N., Shear, P.K., Tripp, L.D., Mayleben, D.W., & Parasuraman, R. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science, 4,* 89-112.

Hollander, T.D., Warm, J.S., Matthews, G., Dember, W.N, Parasuraman, R., Hitchcock, E.M., Beam, C.A., and Tripp, L.D. (2002) Effects of signal regularity and salience on vigilance performance and cerebral hemovelocity. *Proceedings of Human Factors and Ergonomics Society,* 1654-1658.

Mayleben, D.W. (1998). *Cerebral bloodflow velocity during sustained attention.* Unpublished doctoral

dissertation, University of Cincinnati, Cincinnati, OH.

Parasuraman, R., Warm, J.S., See, J.E., (1998). Brain systems of vigilance. In R. Parasuraman, (Ed.), *The attentive brain* (pp. 221-256). Cambridge, MA: MIT Press.

Posner, M.I., (1978). *Chronometric explorations of mind.* Mahwah, NJ: Erlbaum.

Punwani, S., Ordidge, R.J., Cooper, C.E., Amess, P., & Clemence, M. (1998). MRI measurements of cerebral deoxyhaemoglobin concentration [dhB] – correlation with near infrared spectroscopy (NIRS). *NMR in Biomedicine, 11,* 281-289.

Stroobant, N., & Vingerhoets, G. (2000). Transcranial doppler ultrasonography monitoring of cerebral hemodynamics during performance of cognitive tasks: A review. *Neuropsychology Review, 10,* 213-231.

Temple, J.G., Warm, J.S., Dember, W.N., Jones, K.S., LaGrange, C.M., & Matthews, G. (2000) The effects of signal salience and caffeine on performance, workload and stress in an abbreviated vigilance task. *Human Factors, 42,* 183-194.

Toronov, V., Webb, A., Choi, J.H., Wolf, M., Michalos, A., Gratton, E., & Huber, D. (2001). Investigation of human brain hemodynamics by simultaneous near-infrared spectroscopy and functional magnetic resonance imaging. *Medical Physics, 28,* 521-527.

Supported by the Department of Defense MURI Initiative, P.A. Hancock Principal Investigator. Administered by the Army Research Office under grant no. DAAD 19-01-1-062

CHALLENGES TO THE MINDLESSNESS MODEL OF VIGILANCE THROUGH SIGNAL REGULARITY

William S. Helton
Wilmington College
Wilmington, Ohio

Todd D. Hollander, Joel S. Warm, Gerald Matthews, William N. Dember, Matthew Wallaart, Gerald Beauchamp
University of Cincinnati
Cincinnati, Ohio

Raja Parasuraman
Catholic University of America
Washington D.C.

Robertson et al. (1997) have proposed that detection failures in vigilance tasks result from a "mindless" withdrawal of attentional effort from the monitoring assignment. To explore that view, they modified the traditional vigilance task in which observers make button-press responses to signify the detection of rarely occurring critical signals to one in which button-press responses acknowledge frequently occurring non-signal events and response-withholding signifies signal detection. This modification is designed to promote a mindless withdrawal of attentional effort from the task through routinization. The present study challenges the validity of the mindlessness model by showing that with both types of tasks observers can utilize subtle regularities in the temporal structure of critical signal appearances to develop expectations about the time course of those appearances that affect performance efficiency. Such expectations enhance performance with the traditional vigilance task but degrade performance with the modified task.

Vigilance or sustained attention tasks typically require observers to monitor displays over extended periods and to execute overt detection responses to the appearance of low probability critical signals. The signals are usually embedded in a context of recurrent non-signal (neutral) events that, unlike signals, require no overt response from observers. In a recent series of studies, Robertson and his colleagues (Manly, Robertson, Galloway, & Hawkins, 1999; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997) have argued that detection failures in vigilance tasks are due the observers' inattention or 'mindlessness'. According to this perspective, when observers are confronted with the need to respond to infrequent signals separated by long intervals, a supervisory attentional system (Shallice, 1988; Stuss, Shallice, Alexander, & Piction, 1995) loses its potency and observers cease to focus their awareness on the task at hand. Instead, they approach their assignment in a thoughtless, routinized manner characterized by the withdrawal of effortful attention from the task. This approach reflects an endogenous modulation of attention rather than the decline in wakefulness and vigor accompanying lowered arousal (cf. Dickman, 2002).

Based upon their conception, Robertson and his associates introduced a modification to the standard vigilance paradigm designed to more rapidly elicit this mindless state: monitors are required to respond *overtly to non-signals* and to *withhold overt responding to signals*. Presumably, increased detection failures in the modified vigil can be attributed to routinization, automaticity, and lapses of attentional focus, all features of "mindlessness" (Laberge, 1995; Langer, 1989) generated by uniform, repetitive responding to more numerous non-signal events.

When given the opportunity, observers rate traditional vigilance tasks as both mentally demanding and stressful (Helton, Dember, Warm, & Matthews, 1999; Warm, Dember, & Hancock, 1996). A recent study replicated these findings with Robertson's modified task (Grier et al., in press). Contrary to the mindlessness model proposed by Robertson and his colleagues, these results support a resource model view of attention, wherein vigilance tasks are construed as attentionally demanding, thereby exhausting information-processing resources (Matthews, Davies, Westerman, & Stammers, 2000).

While subjective reports of high workload and stress provide evidence against the role of mindlessness

in the standard vigilance paradigm and in the Robertson modification, they do so indirectly. The present study was designed to provide a more direct test of that issue, particularly in regard to the Robertson modification. As described by Coren, Ward, and Enns (1999), observers in attention demanding situations have the ability to actively utilize past experience to form expectancies about anticipated stimulus events and prepare for their occurrence by aligning attention with their anticipated time of arrival. This effect is exemplified in the standard vigilance paradigm by *the signal regularity effect,* wherein performance efficiency is enhanced when the temporal intervals between critical signals for detection occur in a regular and predictable fashion as opposed to an irregular and unpredictable one (Warm, Dember, Murphy, & Dittmar, 1992; Warm, & Jerison, 1984). The ability of observers to be responsive to signal predictability in the case of the Robertson paradigm would constitute strong evidence against a "mindless" approach to the modified vigilance task as well as the standard task.

Method

Twenty observers (10 males and 10 females) were assigned at random to each of four conditions resulting from the factorial combination of task type (standard and modified) and regularity (regular signal schedule and irregular signal schedule). All observers participated in a 40-min vigil divided into four continuous 10-min periods. They inspected the repetitive presentation on a VDT of a simulated air-traffic control display described by Hitchcock et al. (2003). The display comprised a "city" (a solid red circle, 10.5 mm in diameter, luminance = 23.7 cd/m^2) banded by a thin white border (0.75 mm thick x 12 mm in diameter) ringed by three circular white "outer markers" (0.75 mm thick; 28mm, 53 mm, and 83 mm in diameter, respectively; luminance = 79.2 cd/m^2) and two "jet aircraft" (represented by two 1mm x 25 mm lines), all of which were presented on a light gray background (luminance = 29.6 cd/m^2). In all conditions, the display was updated 30 times/min with a dwell time of 300 msec. The Michaelson contrast ratio ([maximum luminance - minimum luminance / maximum luminance + minimum luminance] x 100; Coren, Ward, & Enns, 1999) of the aircraft to their background was 2%. The aircraft were equidistant from the city (each reached the innermost marker), approaching it from opposite headings, either from Northwest to Southeast or Northeast to Southwest. Critical signals for detection (emergency events) were cases in which the two aircraft were aligned on a collision path over the center of the city. Neutral events (safe events) were one of eight permutations of non-collision flight headings. In all conditions, twenty signals were presented per 10-min period of watch. In the *regular signal* condition, inter-signal intervals were fixed at one critical signal every 30 sec; in the irregular signal condition, inter-signal intervals ranged from 12 to 60 sec, with a mean of 30 sec. Stimulus presentations were orchestrated by a MacIntosh personal computer. In the standard condition, participants responded to critical signals by pressing a key on a response pad, whereas they made no overt response to neutral signals. In the modified condition, these instructions were reversed: observers pressed a key in response to neutral events and were to withhold overt responding to critical signals.

Results

Signal detections. Mean percentages of correct detections in all experimental conditions are presented in Figure 1 (error bars are standard errors). In the standard vigil, correct detections were defined as *key presses* to the occurrence of emergency events on the air-traffic control display. In the modified vigil, correct detections were defined as the *withholding of key presses* to such events.

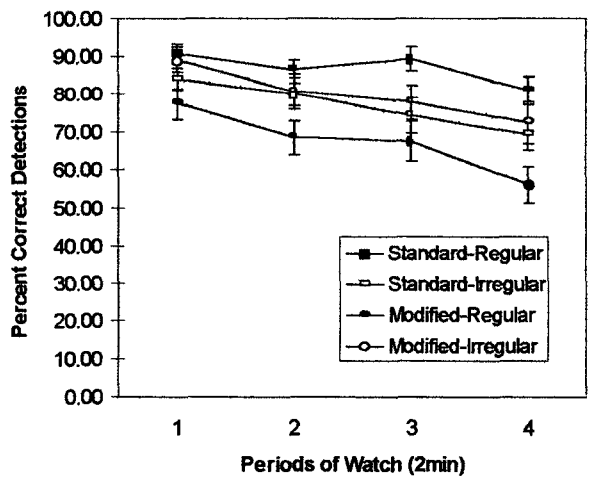


Figure 1. Mean percentages of correct detections in all experimental conditions (error bars are standard errors).

A 2 (task) x 2 (regularity) x 4 (periods of watch) mixed analysis of variance (ANOVA) based upon an arcsin transformation of the detection scores revealed that detection rates in the standard vigil (M = 81.93) were higher than those in the modified vigil (M = 73.78), F (1, 76) = 5.83, $p < .05$, and that the overall level of signal detections declined significantly over time F (3, 218) = 19.36, $p < .001$. Moreover, there was a significant Task x Regularity interaction, F (1, 76) = 12.85 $p < .001$. All other sources of variance in this analysis were non-significant ($p > .05$). In this and all subsequent analyses, Box's epsilon was used when appropriate in computing

degrees of freedom for the repeated measures factors to correct for violations of the sphericity assumption (Maxwell & Delaney 1990).

The Task x Regularity interaction is shown in Figure 2 (error bars are standard errors). The percentage of correct detections was higher for regular ($M = 86.95$) than for irregular ($M = 76.91$) signals in the standard vigil. However, in the modified vigil, the level of signal detections was higher for irregular ($M = 80.09$) than for regular ($M = 67.48$) signals. The regular/irregular signal difference was statistically significant with both types of vigilance tasks, $t(38) > 2.36$, $p < .025$ in each task.

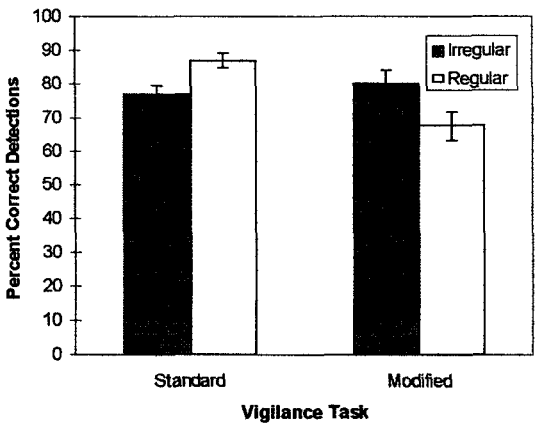


Figure 2. Mean percentages of correct detection for the two vigilance tasks (standard and modified) under the regular and irregular signal conditions (error bars are standard errors).

False Alarms. In the standard vigil, false alarms were defined as *key presses* to the occurrence of safe events on the air-traffic control display. In the modified vigil, false alarms were defined as the *withholding of key presses* to such events. A 2 (task) x 2 (regularity) x 4 (periods of watch) mixed analysis of variance (ANOVA) based upon an arcsin transformation of the false alarm percentages revealed that false alarms were significantly fewer when signals appeared on a temporally regular ($M = 3.73$) as compared to an irregular ($M = 5.85$) schedule, $F(1, 76) = 6.41$, $p < .01$, and that the overall level of false alarms increased over time, $F(2, 176) = 4.87$, $p < .01$. Moreover, this increase over time was dependent upon task, $F(2, 176) = 4.07$, $p < .01$. All other sources of variance in this analysis were non-significant ($p > .05$). The Task x Periods of watch interaction is shown in Figure 3 (error bars are standard errors). It is evident in the figure that while the false alarms rates in the two tasks were similar initially, the rate of gain in false alarms over time was greater in the modified than in the standard vigil. Thus, in the final two periods of watch, false alarms were more prevalent in the modified than in the standard vigil, $t(78) > 2.11$, $p < .05$ in each case.

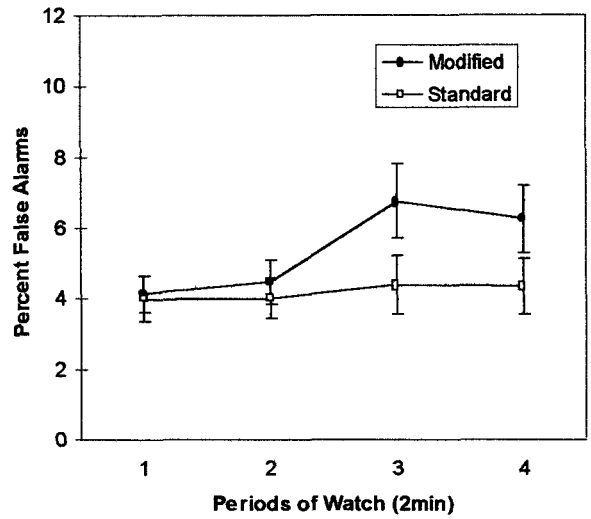


Figure 3. Mean percentages of false alarms for the two vigilance tasks (standard and modified) as a function of periods of watch (error bars are standard errors).

Discussion

Clearly, the results of this study do not support the mindlessness model of vigilance proposed by Robertson and his colleagues (Manly et al., 1999; Robertson et al., 1997). Consistent with prior findings based on the standard vigilance paradigm (Warm et al., 1992; Warm & Jerison, 1884), observers operating under this paradigm detected signals more frequently and made fewer false alarms when the signals appeared on a temporally regular as compared to a temporally irregular schedule. For signal regularity to facilitate performance in this way, the observers needed to be aware of the temporal structure of signal occurrences. Hence, they could not be inattentive or mindless during the task.

The presence of signal regularity also had an impact on performance in the modified format. As in the case of the standard vigil, signal regularity lowered the false alarm rate in the modified vigil but contrary to the finding with the standard vigil, it also lowered the signal detection rate in the modified vigil. The fact that signal regularity affected performance efficiency at all in the modified vigil renders suspect the mindlessness hypothesis in regard to that task because it implies that observers in the modified vigil, like those in the standard vigil, were indeed processing the temporal structure of signal appearances during the watch. The finding that signal regularity lowered signal detections in this type of vigilance task is unusual; it is the first recorded instance of such an outcome with variations in signal regularity and may be the result of conflict between a supervisory

attentional-control system and a task-induced motor routine. Essentially, observers in the modified vigil were confronted with a situation in which they were to make key presses to a frequently occurring imperative stimulus signifying safe events on the air-traffic control display and to make the "response of not responding" when the imperative stimulus occasionally indicated an emergency event. As described by Doyon, Prenhune, and Ungerleider (2003), the need for repetitive key pressing to safe events in the modified vigil is the sort of task requirement that leads to the development of a feed-forward motor scheme, and as Reason (1984) has indicated, well rehearsed motor routines can conflict with the supervisory activities of an attentional-control system. Hence, in the modified condition, the development of expectancies about the temporal appearance of signals could have exaggerated this conflict leading observers in that condition to be less able than their cohorts in the irregular signal condition to execute the proper *response-withholding* action in the presence of critical signals.

While an inability to withhold overt responding when necessary characterizes the performance of observers in the modified vigil when viewed in the context of signal regularity, the opposite effect characterizes their performance when it is viewed in a different context -- in terms of the overall level of false alarms. The higher false alarm rate late in the session of observers in the modified vigil relative to those in the standard vigil implies that the observers in the former condition generally withheld key pressing to safe events more frequently than their cohorts in the standard vigil executed key presses to those events. This effect was also noted by Grier et al. (in press) who suggested that competition between antagonistic action modules (excitation/inhibition; cf. Logan and Cowan, 1984) could have led observers in the modified vigil to have difficulty in initiating the motor response signifying non-signal events. Evidently, the Robertson modification to the standard vigilance paradigm has the potential of eliciting conflict at different levels of perceptual-motor activity. Viewed in this way, the dynamics of the modified vigil transcend the routinized mindlessness envisioned by Robertson and his associates.

References

Coren, S., Ward, L.M., & Enns, J.T. (1999). *Sensation and perception (5th ed.).* Fort Worth, TX: Harcourt-Brace.

Davies, D.R., & Parasuraman, R. (1982). *The psychology of vigilance.* London: Academic Press.

Dickman, S.J. (2002). Dimensions of arousal: Wakefulness and vigor. *Human Factors, 44,* 429-442.

Doyon, J., Penhune, V., Ungerleider, L.G. (2003). Distinct contribution of the cortico-striatal and cortico-cereballar systems in motor skill learning. *Neuropsychologica, 41,* 252-262.

Grier, R.A., Warm, J.S., Dember, W.N., Matthews, G., Galinsky, T.L., Szalma, J.L., & Parasuraman, R. (in press). The vigilance decrement reflects limitations in effortful attention not mindlessness. *Human Factors*

Helton, W. S., Dember, W. N., Warms, J. S., & Matthews, G. (2000). Optimism, pessimism, and false failure feedback: Effects on vigilance performance. *Current Psychology, 18,* 311-325.

Hitchcock, E.M., Warm, J.S., Matthews, G., Dember, W.N., Shear, P.K., Tripp, L.D., Mayleben, D.W., & Parasuraman, R. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomic Science, 4, 89-112.*

LaBerge, D. (1995). *Attentional processing: The brain's art of mindfulness.* Cambridge: Harvard University Press.

Langer, E.J. (1989). *Mindfulness.* Reading, MA: Addison-Wesley.

Logan, G.D., & Cowen, W.B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review, 91,* 295-327.

Manly, T., Robertson, I.H., Galloway, M., & Hawkins, K. (1999). The absent mind: further investigations of sustained attention to response. *Neuropsychologica, 37,* 661-670.

Matthews, G., Davies, D.R., Westerman, S.J., & Stammers, R.B. (2000). *Human performance: Cognition, stress and individual differences.* East Sussex, UK: Psychology Press.

Maxwell, S.E., & Delaney, H.D. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

Reason, J. (1984). Lapses in attention in everyday life. In R. Parasuraman & D.R. Davies (Eds.), *Varities of attention* (pp. 515-549). Orlando, FL: Academic Press.

Robertson, I.H., Manly, T., Andrade, J., Baddeley, B.T., & Yiend, J. (1997). "Oops!": Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologica, 35,* 747-758.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, UK: Cambridge University Press.

Stuss, D.T., Shallice, T., Alexander, M.P., & Picton, T.W. (1995). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences, 769,* 191-209.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In R. Parasuraman and M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183-200). Mahwah, NJ: Prentice-Hall.

Warm, J.S., Dember, W.N., Murphy, A.Z., & Dittmar, M.L. (1992). Sensing and decision-making components of the signal-regularity effect in vigilance performance. *Bulletin of the Psychonomic Society, 30,* 297-300.

Warm, J.S., & Jerison, H.J. (1984). The psychophysics of vigilance (pp. 15-59). Chichester, UK: Wiley.

Supported by the Department of Defense MUIRI Initiative, P.A. Hancock Principal Investigator. Administered by the Army Research Office under grant no. DAAD 19-01-1-062

Accepted for presentation at the meeting of the Southern Society for Philosophy and Psychology, Atlanta GA, April, 2002. Supported by a MURI subcontract from UCF to the University of Cincinnati

Effects of Transitions in Signal Salience on Vigilance Performance and Stress

William S. Helton

Wilmington College

Tyler H. Shaw, Joel S. Warm, Gerald Matthews, William N. Dember,

University of Cincinnati

Peter A. Hancock

University of Central Florida

Laboratory studies of vigilance or sustained attention traditionally maintain a constant information-processing load throughout the experimental session (Davies & Parasuraman, 1982). However, the vigilance tasks encountered in many operational settings, such as air-traffic control, process control, and medical monitoring, can contain abrupt changes in the demands placed upon operators during a duty cycle. Consequently, the National Research Council (Huey & Wickens, 1993) has identified transitions in task demand as an important dimension for study in vigilance research.

Only two published investigations have focused upon this issue. One of these made use of changes in background event rate or the rate of repetition of stimulus events that need to be scanned in order to detect signals. Performance efficiency typically varies inversely with event rate (Warm & Jerison, 1984). Krulewitz, Warm, and Wohl (1975) found that observers shifted abruptly from a low to a high event rate during a vigil did more poorly than non-shifted controls on the high event rate. Those shifted in the high to low direction exceeded the performance of non-shifted controls on the low event rate. An outcome of this sort suggests that the effects of transitions in task demand are characterized by simple psychophysical contrast. However, a subsequent study by Gluckman, Warm, Dember, and Rosa, (1993) using two different vigilance tasks to shift observers from single-task to dual-task monitoring and vice versa produced performance changes that were devoid of contrast effects. In that study, the post-shift performances of the shifted groups equaled that of their non-shifted controls.

Gluckman et al. (1993) suggested that single-task and dual-task conditions in their study may have been perceived as so qualitatively different as to preclude any kind of direct contrast between them. Thus, contrast effects in demand transitions may be limited to changes in a single common dimension. In addition to event rate, performance efficiency in vigilance varies directly

with signal salience (Warm & Jerison, 1984). Accordingly, this study was designed to determine whether contrast effects could be obtained when signal salience was shifted in high-to-low and low-to-high directions.

To date, studies of demand transitions have only examined effects on performance efficiency. However, vigilance tasks are stress-inducing, and the stress of sustained attention is closely tied to psychophysical demands (Galinsky, Warm, Dember, & Rosa, 1968; Hancock & Warm, 1987; Temple et al., 2000). Hence, a second goal for this study was to examine the effect of transitions in task demand on task-induced stress in vigilance.

Twenty observers (10 males and 10 females) were assigned at random to each of four conditions resulting from the factorial combination of signal salience (high and low salience signals) and switching (switch and no-switch). All observers participated in a 12-min vigil divided into six continuous 2-min periods during which they were required to detect a target letter in a visual noise field. This abbreviated vigilance task duplicates the general effects of signal salience and task-induced stress noted with more traditional long-duration tasks (Temple et al., 2000). Switch participants experienced 6-min at one salience level and then 6-min at the other. Perceived stress was assessed by means of the Dundee Stress State Questionnaire (DSSQ; Matthews, Joyner, Gilliland, Huggins, & Falconer, 1999), which provides factor-analytically differentiated scales of Task Engagement, Distress, and Worry.

Consistent with the findings of Gluckman et al. (1993), the post-shift performance of shifted observers in this study simply equaled that of their non-shifted controls. However, shifted observers were significantly more distressed at the end of the vigil than non-shifted participants. Thus, even when shifts in task demand do not produce differential performance effects, they can have differential consequences for the stress of sustained attention.

References

Davies, D.R., & Parasuraman, R. (1982). *The psychology of vigilance.* London: Academic Press.

Galinsky, T.L., Rosa, R.R., Warm, J.S., & Dember, W.N. (1993). Psychophysical determinants of stress in sustained attention. *Human Factors, 35,* 603-614.

Gluckman, J.P., Warm, J.S., Dember, W.N., & Rosa, R.R. (1993). Demand transitions and sustained attention. *The Journal of General Psychology, 120,* 323-337.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519-537.

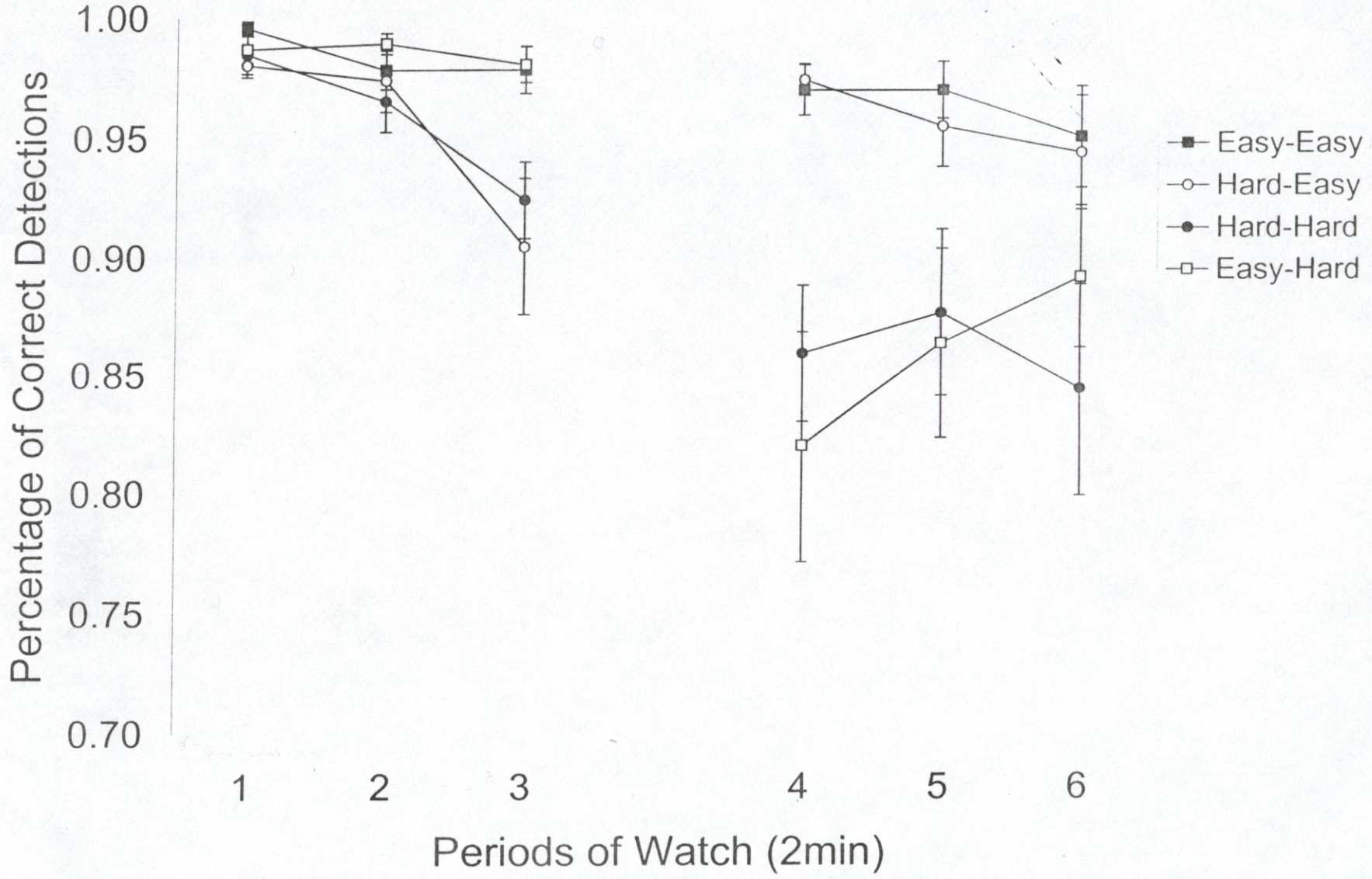
Huey, B.M. & Wickens, C.D. (Eds.) (1993). *Workload transitions: Implications for individual and team performance.* Washington D.C.: National Academy Press.

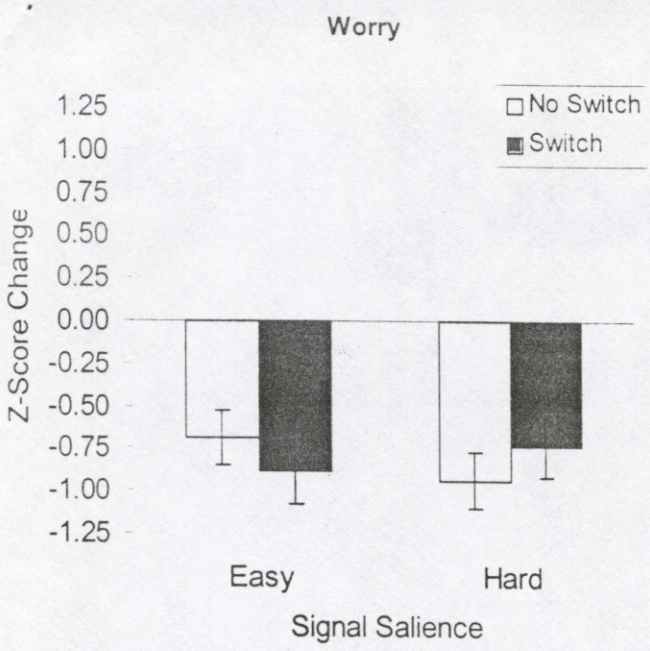
Krulewitz, J.E., Warm, J.S., & Wohl, T.H. (1975). Effects of shifts in the rate of repetitive stimulation on sustained attention. *Perception & Psychophysics, 18,* 245-249.

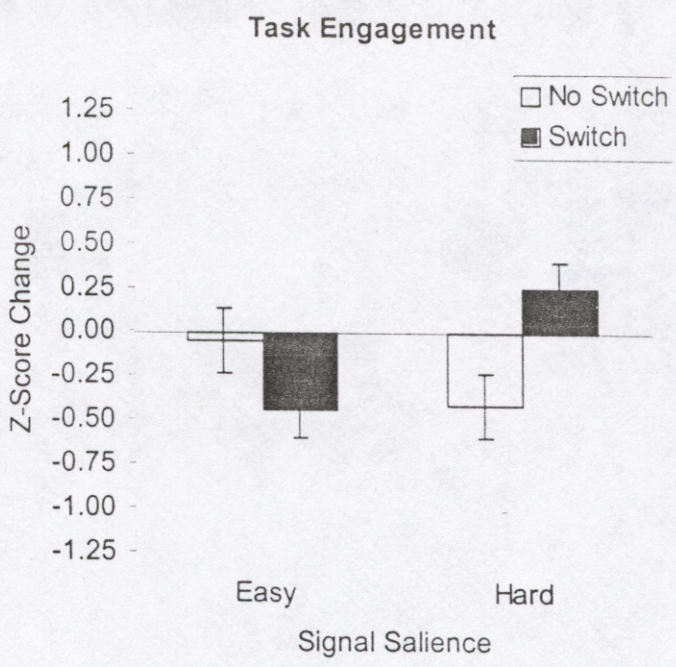
Matthews, G., Joyner, L., Gilliland, K., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a big state three? In I. Merville, I.J. Deary, F. DeFruyt, and F. Ostendorf (Eds.), *Personality psychology in Europe* (vol. 7) pp. 335-350. Tilburg: Tilburg University Press.

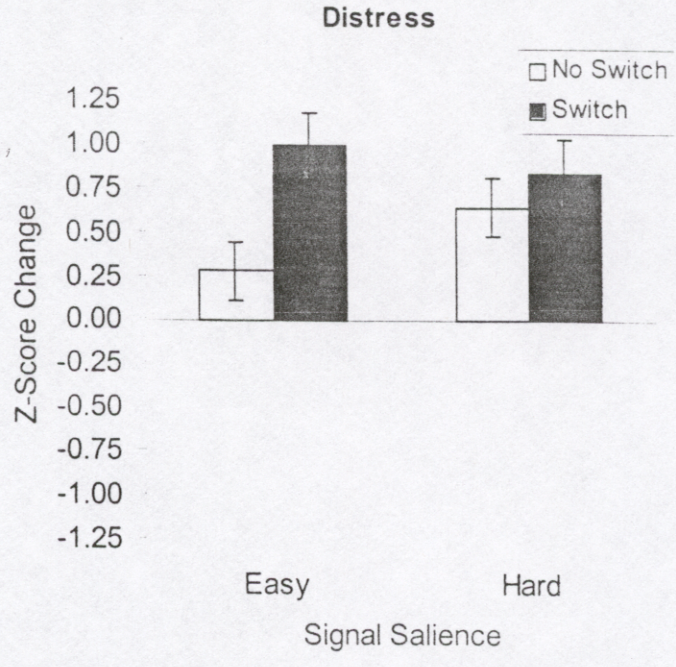
Temple, J.G., Warm, J.S., Dember, W.N., Jones, K.S., LaGrange, C.M., & Matthews, G. (2000) The effects of signal salience and caffeine on performance, workload and stress in an abbreviated vigilance task. *Human Factors, 42,* 183-194.

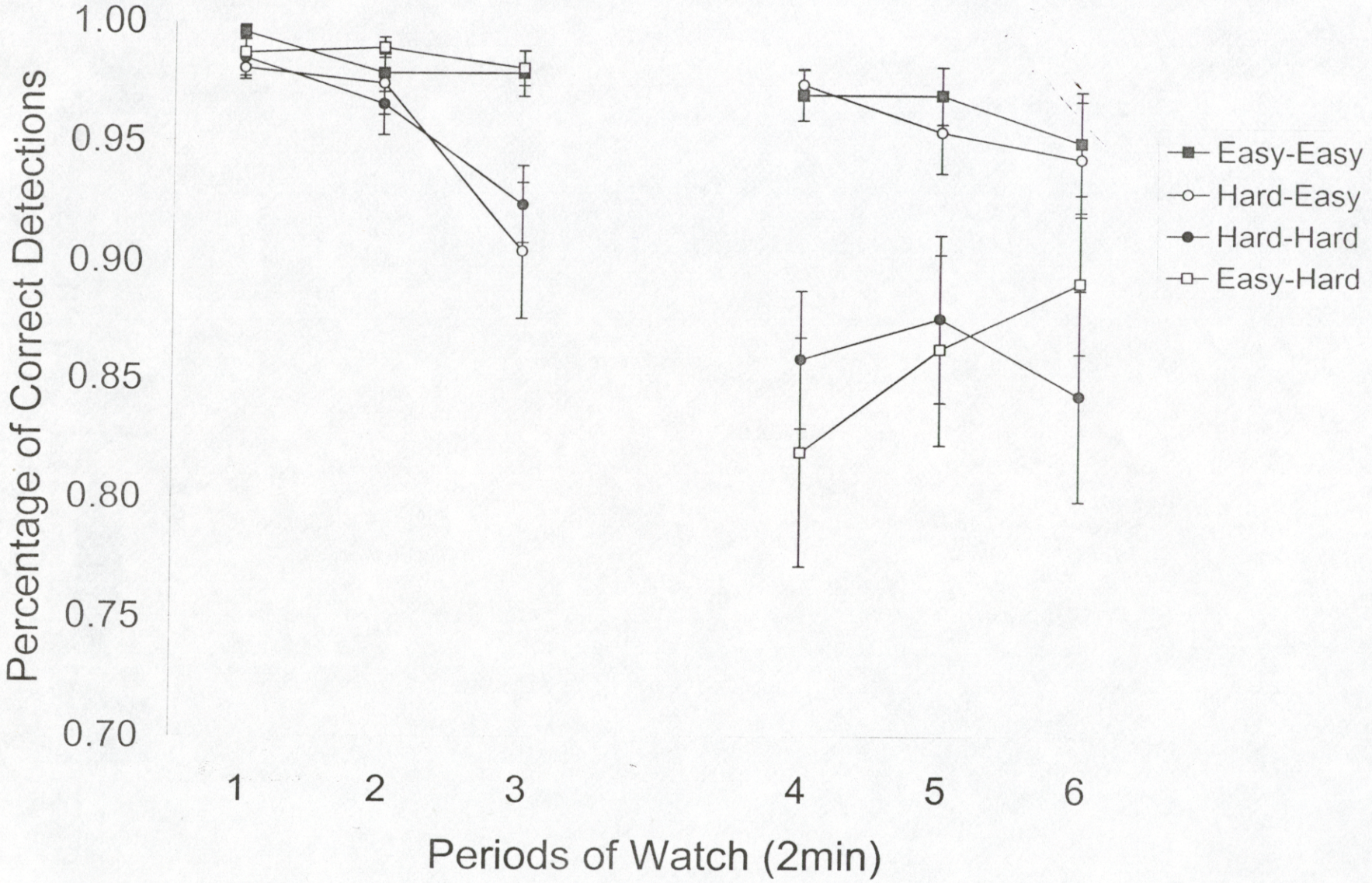
Warm, J.S., & Jerison, H.J. (1984). The psychophysics of vigilance. In J.S. Warm (Ed.), *Sustained attention in human performance* (pp. 15-59). Chichester, UK: Wiley.

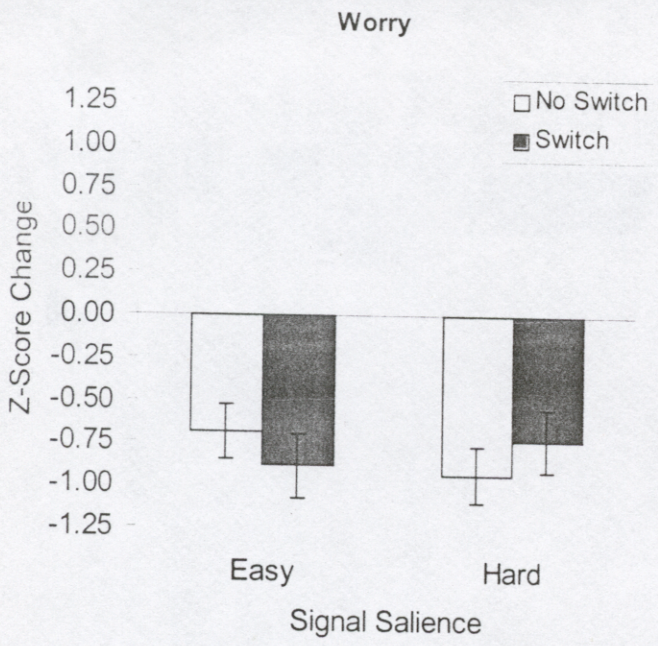

1.00
0.95
0.90
0.85
0.80
0.75
0.70
Percentage of Correct Detections
1
2
3
4
5
6
Periods of Watch (2min)
Easy-Easy
Hard-Easy
Hard-Hard
Easy-Hard

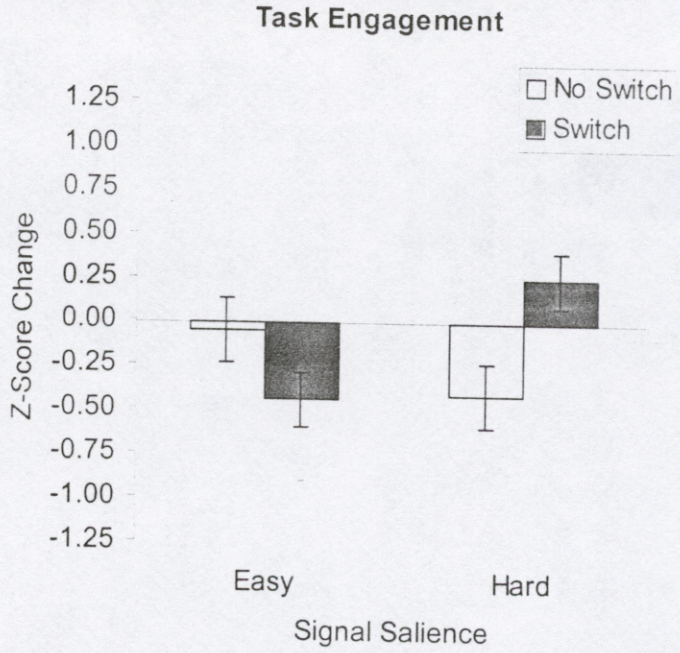

Worry
No Switch
Switch
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
Z-Score Change
Easy
Hard
Signal Salience

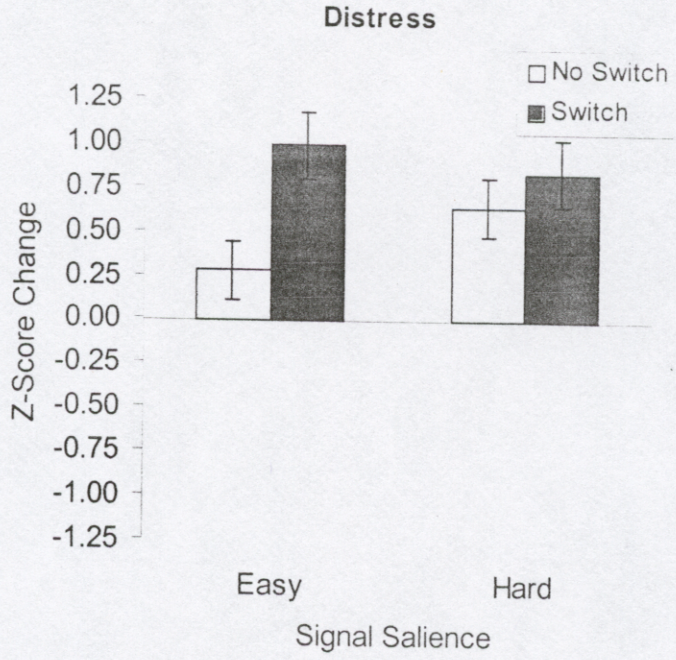

Task Engagement
No Switch
Switch
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
Z-Score Change
Easy
Hard
Signal Salience


Distress
No Switch
Switch
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
Z-Score Change
Easy
Hard
Signal Salience


1.00
0.95
0.90
0.85
0.80
0.75
0.70
Percentage of Correct Detections
1
2
3
4
5
6
Periods of Watch (2min)
Easy-Easy
Hard-Easy
Hard-Hard
Easy-Hard


Worry
No Switch
Switch
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
Z-Score Change
Easy
Hard
Signal Salience


Task Engagement
No Switch
Switch
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
Z-Score Change
Easy
Hard
Signal Salience


Distress
No Switch
Switch
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
Z-Score Change
Easy
Hard
Signal Salience

Quarterly Progress Report
Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield
Army MURI, Subcontract (*Models of Workload Transition*) to U. of Central Florida

Gerald Matthews and Joel S. Warm
University of Cincinnati

Performance Period: 1/1/03-3/31/03

During this period, data analysis and write-ups were completed for three studies to be presented at the annual meetings of the Southern Society for Philosophy and Psychology, Atlanta, GA and the Human Factors and Ergonomics Society in Boulder, CO. In addition, Drs. Matthews and Szalma organized a symposium on stress at the Southern Society meeting and Dr. Matthews was appointed program chair for the 100th anniversary meeting of the Southern Society to be held in New Orleans, LA in April 2004.

Paper Presentations

1.*William S. Helton, Tyler S. Shaw, Joel S. Warm, Gerald Matthews, William N. Dember,& Peter A. Hancock,* Effects of Transitions in Signal Salience on Vigilance Performance and Stress. Paper presented at the annual meeting of the Southern Society for Philosophy, Atlanta GA, April 17-19, 2003.

The National Research Council (Huey & Wickens, 1993) has identified transitions in task demand as an important dimension for consideration in military and civilian applications. This study focuses upon workload transitions in regard to vigilance performance. The effects of such transitions have been explored in two earlier studies from our laboratory. The results of the first study (Krulewitz, Warm, & Wohl, 1975) suggested that the effects of transitions in task demand are characterized by simple psychophysical contrast. However, a second study (Gluckman, Warm, Dember, & Rosa, 1993) failed to confirm these results. The present study was designed to further explore whether a psychophysical contrast model can adequately describe demand transitions. With that goal in mind, observers were shifted from a high salience (low demand) to a low salience (high demand) condition and vice-versa while performing the abbreviated vigilance task described in our last quarterly report. Control observers were maintained at either high or low salience levels throughout the watch. The results failed to confirm the psychophysical contrast model. Consistent with the findings of Gluckman et al. (1993) the post-shift performance of observers in this study simply equaled that of their non-shifted controls. Although shifts in demand levels had no effect upon performance efficiency, they had a notable effect upon observers' self-reports of stress as indexed by the DSSQ multidimensional stress state questionnaire (Matthews, et al., 1999; 2002). Regardless of the direction of shift in task demand, shifted observers were significantly more distressed at the end of the vigil than non-shifted controls. In addition, when observers were shifted from a low to a high salience condition, task engagement declined, whereas when they were shifted from a high to a low salience condition, task engagement increased. The presence of distress-engagement dissociation in the present data provides further evidence for the utility of the multi-dimensional view of stress advocated by Matthews and his associates (1999; 2002). The findings that transitions in task demand can produce increased feelings of distress, and in the case of shifts in the high-to-low demand direction (low to high salience), can also decrease task engagement may have long-term implications for participant well-being in operational settings.

2. *Joel S. Warm, Lloyd D. Tripp, Kelley S. Parsons, Todd D. Hollander, Rhiannon Hamilton, Gerald Matthews, William N. Dember, William S. Helton, and Raja Parasuraman,* Cerebral Hemodynamics and Performance on an Abbreviated Vigilance Task. Paper presented at the meeting of the Southern Society for Philosophy and Psychology, Atlanta, GA, April 17-19 2003. An expanded version of this paper has been accepted for presentation at the annual meeting of the Human Factors and Ergonomics Society, Denver, CO, October 13-17, 2003.

Transcranial Doppler sonography (TCD) and transcranial oximetry (TCCO) measures of blood flow and oxygenization levels in the brain were collected while observers performed our abbreviated 12

min vigilance task that is designed to serve as an analog to more traditional long-duration vigils. Both measures showed higher levels of vascular activity in the right as compared to the left cerebral hemisphere indicating that the overall level of performance in the abbreviated vigil is right-lateralized, a finding that coincides with the outcome of earlier blood flow studies featuring more traditional long-duration vigils. This parallel provides strong support for the argument that the abbreviated vigil is a valid analog of more traditional long-duration vigilance tasks and implies that laterality in vigilance is a generalized effect in terms of cerebral vascular dynamics that appears in terms of both blood oxygenation and hemovelocity.

3. *William S. Helton, Todd D. Hollander, Joel S. Warm, Gerald Matthews, William N. Dember, Matthew Wallaart, Gerald Beauchamp, & Raja Parasuraman,* Challenges to the Mindlessness Model of Vigilance Through Signal Regularity. Paper accepted for presentation at the annual meeting of the Human Factors and Ergonomics Society, Denver CO, October 13-17, 2003.

Robertson et al. (1997) have proposed that detection failures in vigilance tasks result from a "mindless" withdrawal of attentional effort from the monitoring assignment. To explore that view, they modified the traditional vigilance task in which observers make button-press responses to signify the detection of rarely occurring critical signals to one in which button-press responses acknowledge frequently occurring non-signal events and response-withholding signifies signal detection. This modification is designed to promote a mindless withdrawal of attentional effort from the task through routinization. Previous work from our lab (Grier et al. in press) challenges the validity of the mindlessness model by showing that the workload and stress levels associated with the traditional and modified vigilance formats are equally high. This study provides an additional challenge to the validity of the mindlessness model by by showing that with both types of formats, observers performing our simulated air-traffic control display can utilize subtle regularities in the temporal structure of critical signal appearances to develop expectations about the time course of those appearances that affect performance efficiency. Such expectations enhance performance with the traditional vigilance format but degrade performance with the modified format. In either case, reaction to subtle temporal regularities in the temporal structure of critical signal appearances is contrary to the thoughtless, routinized approach to the vigilance task depicted in the "mindlessness" model.

Southern Society for Philosopy and Psychology Symposium

The symposium organized by Drs. Matthews and Szalma focused upon developing a cognitive theory of stress and performance. As part of that symposium, Dr. Matthews presented a paper describing how changes in stress states mediate between external stressors and performance. He illustrated this effect using a study conducted in our lab on the effects of colds on vigilance performance in which a regression analysis demonstrated that stress state could predict the degrading effects of colds on signal detection. The title of that presentation was:

Gerald Matthews, Joel S. Warm, & William S. Helton, Mediators of Stress Effects: Subjective States and Cognition.

Ongoing Studies

1. Work on cerebral vascularity and vigilance performance is continuing. The Warm et al. study described above was the initial examination of the relation between transcranial cerebral oximetry (TCCO) and vigilance performance. Using the abbreviated vigil, that study did not find that the vigilance decrement was paralleled by a decline in the TCCO measure nor with the TCD measure of cerebral hemovelocity. It is conceivable that the abbreviated vigil was too brief to detect temporal changes in cerebral hemovelocity for either measure. To explore this issue more fully, a study is being conducted using the TCCO measure with the simulated air-traffic control task, a more traditional long-duration vigilance task in which the vigilance decrement has been accompanied by declines in the TDC measure of cerebral hemocvelocity (Hitchcock, et al. 2003).

2. Studies by Matthews and Davies (1998) have demonstrated that increments in Energetic Arousal enhance the availability of information processing resources and enable observers to counter the negative effects of increments in psychophysical demand upon vigilance performance. A study is in progress to extend this line of investigation using a particularly severe psychophysical challenge, increments in

background event rate. This study also employs the TCCO measure to determine the relation between blood oxygenation and energetic arousal.

3. Prior work from our lab has centered upon workload transition in vigilance performance. A study is being developed to extend this line of investigation to tracking performance. The aim for the study is to test structural and strategic models of the effects of workload history on performance efficiency. Two levels of tracking difficulty, easy and hard, will be combined with single-task (tracking alone) and dual-task (tracking plus vigilance) conditions. Transition from high to low workload will be accomplished by switching from the dual-task (tracking and vigilance) to the single-task (tracking alone) condition. The structural model or the *resource depletion model,* maintains that an initially high workload will consume information-processing resources that are not replenished. Therefore, needed resources will not be available when participants transition to a low workload situation. Consequently, this model leads to the expectation of performance impairment on the single-tracking task following transition from the dual-task, particularly when the tracking task is more difficult. The strategic model or, *the effort mobilization model,* focuses upon participants' attitudes toward the task and the effort that they put forth in coping with the task. According to this view, effort mobilization depends upon tasks demands and the perceived contrast in demand when shifting from a dual-task to single-task condition will lead to effort reduction in performing the single task, particularly when the tracking task is easy. Consequently, this model leads to the expectation of performance impairment on the single task following transition from the dual-task and that such impairment will be maximal when the tracking task is easy. Instrumentation on this project has been accomplished and pilot testing is under way to determine the appropriate difficulty levels of the tracking task and the nature of the vigilance task to be employed in the dual-task condition. Data collection on this project will begin in June. In addition to performance measures, The NASA-TLX and DSSQ scales will be used to measure workload and stress, respectively, and the Coping Inventory for Task Situations (CITS, Matthews & Campbell, 1998) will be used to assess coping strategies. In addition, the TCD procedure will be employed to provide a metabolic index of resource utilization.

Quarterly Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield

Army MURI, Subcontract (*Models of Workload Transition*) to U. of Central Florida

Gerald Matthews and Joel S. Warm

University of Cincinnati

Performance Period: 4/1/03-5/31/03

During this period, the main focus of research was on programming and piloting for a study of workload transition from dual-task to single-task performance. In addition, Drs. Matthews and Warm participated in a U.S. Army workshop on Cognitive, Psychophysiological, and Behavioral Monitoring for Military Field Applications workshop in Covington, KY, and a steering committee meeting for the U.S. Army Medical Research and Materiel Command program effort (Science and Technology Package, STP C). At this meeting, held from June 10-12, workload transition was identified as a focal area for future work under this program.

Ongoing Studies

1. A task is being programmed and piloted, for use in a forthcoming empirical study of workload transition in vigilance performance. As described in the previous report (1/1/03-3/31/03), the study aims to test two different models of the effects of workload history on performance efficiency, the resource depletion model and the effort-mobilization model. A first-order tracking task has been programmed in Visual Basic that requires the participant to maintain a cross stimulus within a box in the center of the screen. Lateral position of the stimulus varies unpredictably according to a forcing function (the sum of two sine waves of different periods). A joystick control allows the participant to control lateral position, countering deviations from the central position.

Testing the two models of workload transition requires manipulation of workload of the task so as to investigate how varying task demands influences its sensitivity to performance deficits resulting from the transition from dual- to single-task performance. Thus, careful task design is necessary in order to obtain valid tests of the predictions made by these models. Piloting to date has been directed towards development of high- and low-workload versions of the task, using both the objective index of root mean square error in performance, and the subjective index provided by the NASA-TLX workload scale. Several techniques for varying difficulty, including changes in periodicity and gain, were explored. Six pilot subjects completed up to four different versions of the task, working for 20 minutes on each version. These pilot studies have successfully identified two versions of the task, differing in both subjective and objective workload, that will be used in the subsequent study. The pilot studies also showed that practice effects were limited to the first five minutes of the tasks, data that will inform the protocol for practice in the main study.

The study will investigate the transition from dual-task (tracking together with vigilant monitoring for signals) to single-task (tracking alone) performance. Thus, the next stage of task design is to embed a vigilance task into the tracking task, by changing the stimulus attributes of the target that is tracked. Programming of some candidate tasks is in progress, and we will soon initiate piloting of these tasks, alone and in conjunction with the tracking task.

Quarterly Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield

Army MURI, Subcontract (*Models of Workload Transition*) to U. of Central Florida

Gerald Matthews and Joel S. Warm

University of Cincinnati

Performance Period: 7/1/2003 - 8/31/2003

During this period, the main focus of research was the continuing piloting for a study of workload transition from dual-task to single-task performance. In addition, Dr. Matthews presented a paper covering some of the empirical studies of stress, workload and attention described in previous progress reports to the International Society for the Study of Individual Differences, held in Graz, Austria, from July 13-17.

Ongoing Studies

As described in the previous report (4/1/2003 - 6/30/2003), a task is being programmed in Visual Basic and piloted for use in a forthcoming empirical study of workload transition in vigilance performance. The study aims to test two different models of the effects of workload history on performance efficiency, the resource depletion model and the effort-mobilization model. The previous report described how we developed task versions that differed in the level of workload imposed by the tracking task, assessed using objective performance indexed by root mean square error, and the NASA Task Load Index (TLX).

Subsequent piloting efforts have been directed towards programming an integrated vigilance task for use in a dual-task paradigm. Several versions of the vigilance task have been investigated in pilot studies. It is apparent that not all vigilance tasks interfere with tracking performance as predicted by the resource model. However, the latest version of the task incorporates a vigilance task that appears to generate dual-task interference. It requires the participant to discriminate dot configurations presented within the cross-hair stimulus used in the tracking task. Task instructions have been modified to place the task within a military context: piloting a helicopter (tracking) while monitoring for hostile aircraft (vigilance).

We have consulted with an outside expert in workload history: Dr. Eugenia Cox-Fuenzalida of the Department of Psychology, University of Oklahoma. The final design for the study incorporates her recommendations. Due to high inter-individual variability in tracking performance revealed by the pilot studies, we will include a period of single-task baseline performance prior to the workload manipulation, allowing workload transition effects to be tested against a within-subjects estimate of error. We aim to begin data collection during the autumn quarter of the academic year. A human factors graduate student, Nat Unger, has been assigned primary responsibility for data collection and management.

Conference presentation

Dr. Matthews presented an overview of research on stress and workload at UC at the ISSID meeting, including data from studies supported by the current grant.

The abstract follows:

Title: Subjective State Scales as Predictors of Performance in Stressful Environments

Authors: Matthews, G., Warm, J.S., & Helton, W.S.

Recent psychometric and experimental studies have identified three broad factors that describe subjective states: task engagement, distress and worry (Matthews et al., in press). For example, task engagement represents an integration of states of energetic arousal, motivation and concentration. Valid state constructs should correlate with objective behavioral measures. This paper summarizes two studies that tested the criterion validity of the Dundee Stress State Questionnaire (DSSQ: Matthews et al., in press), as a predictor of performance on a visual vigilance task. In both studies, an external stressor was manipulated: naturally-occurring infection with the common cold (Study 1; N=204), and exposure to loud (max. 95 dBA) aircraft engine noise (Study 2; N=192). Both stressors influenced signal detection, and DSSQ state factors, especially task engagement, correlated with significantly higher detection rates in both studies. Further analyses showed that stressor effects were fully mediated by the state factors. These results contribute both to validation of the DSSQ and to development of a comprehensive model for individual differences in response to stressors.

Quarterly Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield
Army MURI Subcontract (models of Workload Transition) to U of Central Florida

Gerald Matthews and Joel S. Warm
University of Cincinnati

Performance Period: 9/01/03-10/31/03

During this period, three papers were presented to the annual conference of the Human Factors and Ergonomics Society in Denver, Colorado and Dr. Matthews contributed to a panel session at that meeting on individual differences and human factors research. Dr. Matthews and Dr. Warm also presented a summary of MURI supported research efforts in their laboratory at the University of Cincinnati to the MURI review and planning panel that met at the University of Central Florida. Dr. Warm accepted an invitation to serve as a plenary speaker at the Fifth Automation Technology and Human Performance Conference to be held in Daytona Beach, Florida in April 2004. In addition, Dr. Warm was nominated to serve as a member of the National Academy of Science National Research Council Committee on Human Factors. Four papers were submitted for presentation at future meetings, one manuscript was submitted for publication and four studies are currently in progress.

Paper Presentation

1. *William S. Helton, Todd D. Hollander, Joel S. Warm, Gerald Matthews, William N. Dember, Matthew Wallaart, Gerald Beauchamp and Raja Parasuraman,* Challenges to the Mindlessness Model of Vigilance Through Signal Regularity. Robertson et al. (1997) have proposed that detection failures in vigilance tasks result from a "mindless" withdrawal of attentional effort from the monitoring assignment. To explore that view, they modified the traditional vigilance task in which observers make button-press responses to signify the detection of rarely occurring critical signals to one in which button-press responses acknowledge frequently occurring non-signal events and response withholding signifies signal detection. This modification is designed to promote a mindless withdrawal of attentional effort from the task through routinization. This study challenges the validity of the mindlessness model by showing that with both types of tasks observers utilize subtle regularities in the temporal structure of critical signal appearances to develop expectations about the time course of those appearances that affect performance efficiency. Such expectations enhance performance with the traditional vigilance task but degrade performance with the modified task.

2. *Todd D. Hollander, William S. Helton, Kelly S. Parsons, Joel S. Warm, Gerald Matthews, and William and Dember,* Cerebral Hemodynamics and Performance on an Abbreviated Vigilance Task. Transcranial Doppler sonography (TCD) and transcranial oximetry (TCCO) measures of blood flow and oxygenation levels in the brain were collected while observers performed an abbreviated 12 -min vigilance task developed in our laboratory that is designed to serve as an analog to more traditional long-duration vigils. Both measures showed higher levels of vascular activity in the right as compared to the left cerebral hemisphere, indicating that the overall level of performance in the abbreviated vigil is right-lateralized, a finding that coincides with the outcome of earlier blood flow studies featuring more traditional long-duration vigils. This parallel provides strong support for the argument that the abbreviated vigil is a valid analog of more traditional long-duration vigilance tasks and implies that laterality in vigilance is a generalized effect in terms of cerebral vascular dynamics that appears in terms of both blood oxygenation and hemovelocity.

3. *Gerald Matthews, Amanda Emo*, and *Greg Funke*, Emotional Intelligence: Implications for Human Factors. This paper reviews the relevance to the human factors practitioner of emotional intelligence (EI), the ability to understand and manage emotion. Assessment of EI may be useful for predicting individual differences in performance in stressful operational environments. However, existing research demonstrates some difficulties in developing valid tests for EI. Self-report and objective tests for EI fail to converge well, and may be measuring different constructs. Objective tests are promising, but further research is needed to establish their predictive validity in human factors contexts. Three examples are given of applied contexts

in which assessment of EI may be valuable: performance under stress, coping with anger and frustration and teamwork. Tests for EI might be used either for operator selection, or as an adjunct to training, but further research is needed to validate their use. The paper concludes with some initial guidelines for deciding whether EI is potentially relevant to a particular operational setting.

4. *Gerald Matthews*. Considering the importance of individual differences in human factors research: No longer simply confounding noise (Panel Session). This short presentation aimed to stimulate discussion about the role of theoretical models in practical applications of individual differences research. It was suggested that the person-machine transaction operates at both biocognitive and cognitive-adaptive levels. The biocognitive transaction refers to the effects of the task environment on parameters of the neural and cognitive architectures, such as resource availability, speed of processing and memory capacities, whereas the cognitive-adaptive level refers to the operator's intentions, situation evaluation and strategies for coping with task demands. Biocognitive and cognitive-adaptive biases work together to influence the affordances of the environment for the individual. For example, stress-vulnerable individuals such as those high in general or driving anxiety are adapted for avoiding environmental threat, whereas hardy individuals are better prepared to meet threats directly. At a practical level, differential psychology has two messages for designers. The first is that indices of functional status differ in their validity depending on personality characteristics. The second is that the personal significance of the task environment has profound implications for performance and wellbeing, so that successful design requires an understanding of the meanings that the operator will 'read into' the system.

Paper Submissions

1. *William S. Helton, Tyler S. Shaw, Joel S. Warm, Gerald Matthews, William N. Dember, and Peter A. Hancock,* Effects of Transitions in Signal Salience on Vigilance Performance and Stress. Submitted to the Fifth Automation Technology and Human Performance Conference, Dayton Beach, Florida April 2004. The National Research Council (Huey & Wickens, 1993) has identified transitions in task demand as an important dimension for consideration in military and civilian applications. This study focuses upon workload transitions in regard to vigilance performance. The effects of such transitions have been explored in two earlier studies from our laboratory. The results of the first study (Krulewitz, Warm, & Wohl, 1975) suggested that the effects of transitions in task demand are characterized by simple psychophysical contrast. However, a second study (Gluckman, Warm, Dember, & Rosa, 1993) failed to confirm these results. The present study was designed to further explore whether a psychophysical contrast model can adequately describe demand transitions. With that goal in mind, observers were shifted from a high salience (low demand) to a low salience (high demand) condition and vice-versa while performing the abbreviated vigilance task described in our last quarterly report. Control observers were maintained at either high or low salience levels throughout the watch. The results failed to confirm the psychophysical contrast model. Consistent with the findings of Gluckman et al. (1993), the post-shift performance of observers in this study simply equaled that of their non-shifted controls. Although shifts in demand levels had no effect upon performance efficiency, they did have a notable effect upon observers' self-reports of stress as indexed by the DSSQ multidimensional stress state questionnaire (Matthews et al., 1999; 2002). Regardless of the direction of shift in task demand, shifted observers were significantly more distressed at the end of the vigil than non-shifted controls. In addition, when observers were shifted from a low to a high salience condition, task engagement declined, whereas when they were shifted from a high to a low salience condition, task engagement increased. The presence of distress-engagement dissociation in the present data provides further evidence for the utility of the multi-dimensional view of stress advocated by Matthews and his associates (1999; 2002). The findings that transitions in task demand can produce increased feelings of distress, and in the case of shifts in the high-to-low demand direction (low to high salience), can also decrease task engagement may have long-term implications for participant well-being in operational settings. A prior version of this paper was submitted to the Human Factors and Ergonomic Society meeting in Baltimore, MD, 2002 but was not placed on the program because of an administrative error. If accepted at the automation conference, the study will appear as a chapter in a book describing the proceedings of the conference.

2. *Todd D. Hollander, Joel S. Warm, Gerald Matthews, William N. Dember, Lloyd D. Tripp, and Raja Parasuraman.* Cerebral Hemovelocity and the Signal Regularity Effect in Vigilance. Submitted to the

meeting of the Psychonomic Society, Vancouver British Columbia, November 2003.. The signal regularity effect -- enhanced performance efficiency when critical signals for detection appear in a temporally regular, as opposed to a temporally irregular manner – has a long history in vigilance research. Nevertheless, the precise psychophysical relations under which this effect can be elicited have not been identified. Toward that end, this study demonstrates that the effect is limited to low-salience signals perhaps because the effort needed to generate veridical temporal expectancies is unnecessary with high salience signals. Additionally, using signal detection theory metrics (*d' and c)* and neuroimaging of cerebral blood flow via transcranial Doppler sonography, this study also shows that the signal regularity effect is rooted in sensing rather than decision-making factors and that it is lateralized to the right cerebral hemisphere. This is a more extensive version of an earlier poster presented to the Human Factors and Ergonomic Society, Baltimore, MD. September, 2002.

3. *Kelly S. Parsons, Lloyd D. Tripp, Joel S. Warm, Gerald Matthews, Lauren E. Reinerman, & Raja Parasuraman,* Transcranial Cerebral Oximetry and Vigilance Performance. Submitted to the meeting of the Southern Society for Philosophy and Psychology, New Orleans, LA, April, 2004. An earlier study from our laboratory demonstrated that a measure of frontal lobe oxygen saturation using transcranial cerebral oximetry (TCCO) paralleled the effects of cerebral blood flow reflected via transcranial Doppler sonography (TCD). Both measures demonstrated increased overall activity in the left cerebral hemisphere when observers performed an abbreviated 12-min vigilance task. In that study, neither measure was successful in showing that the vigilance decrement, the decline in signal detections over time that typifies vigilance performance, was accompanied by changes in cerebral hemodynamics. TCD-measured declines in blood flow have been found, however, in other studies using longer work periods. Consequently, the present study made use of a more traditional long-duration (40-min) vigilance task featuring a simulated air-traffic control display in which the vigilance decrement was accompanied by a temporal decline in blood flow (Hitchcock et al., 2003) to determine if the TCCO procedure will reveal comparable temporal declines in cerebral oxygen saturation. The performance decrement in this study was indeed accompanied by a significant temporal decline in cerebral oxygen saturation. Like the prior findings with blood flow, the temporal decline in oxygen saturation only appeared when observers were required to actively perform the vigilance task; the level of oxygen saturation remained stable over time in participants who observed the display without work a imperative. The results indicate that the TCCO procedure can serve as a useful adjunct to the measurement of cerebral blood flow in imaging cerebral hemodynamics during the performance of a vigilance task.

4. *Joel S. Warm,* Cerebral Hemodynamics and Vigilance Performance. Submitted to the meeting of the Southern Society for Philosophy and Psychology, New Orleans, LA. This paper is an invited overview of our work on cerebral hemodynamics (blood flow and oxygen saturation) in vigilance to be presented at a session honoring Dr. Paul S. Siegel, who passed away this year. Dr. Siegel was formerly the Chair of the Department of Psychology of the University of Alabama and a distinguished member of the Southern Society. He was one of Dr. Warm's greatest teachers.

Manuscript Submissions

1. *Helton, W.S., Hollander, T.D., Warm, J.S., Mathews, G., Dember, W.N., Wallart, M., Beauchamp,G., Parasuraman, R., & Hancock, P.A.,* Signal Regularity and the Mindlessness Model of Vigilance, submitted to the *British Journal of Psychology***.** The manuscript is based upon the Helton, et al., talk given to the Human Factors and Ergonomics Society in October 2003 described above.

Ongoing Studies

1. Studies by Matthews and Davies (1998) have demonstrated that increments in Energetic Arousal enhance the availability of information processing resources and enable observers to counter the negative effects of increased psychophysical demand upon vigilance performance. A study is in progress to extend this line of investigation by using a particularly severe psychophysical challenge, increased background event rate, This study also employs the TCCO measure to determine the relation between blood

oxygenation and energetic arousal. The study will serve as a Master's Thesis for a first-year graduate student, *Mr. Victor Finemore*

2. Prior work from our lab has centered upon workload transition in vigilance performance. This study extends this line of investigation to tracking performance. The aim for the study is to test structural and strategic models of the effects of workload history on performance efficiency. Two levels of tracking difficulty, easy and hard, will be combined with single-task (tracking alone) and dual-task (tracking plus vigilance) conditions. Transition from high to low workload will be accomplished by switching from the dual-task (tracking and vigilance) to the single-task (tracking alone) condition. The structural model or the *resource depletion model,* maintains that an initially high workload will consume information-processing resources that are not replenished. Therefore, needed resources will not be available when participants transition to a low workload situation. Consequently, this model leads to the expectation of performance impairment on the single-tracking task following transition from the dual-task, particularly when the tracking task is more difficult. The strategic model or *the effort mobilization model,* focuses upon participants' attitudes toward the task and the effort that they put forth in coping with the task. According to this view, effort mobilization depends upon tasks demands and the perceived contrast in demand when shifting from a dual-task to single-task condition will lead to effort reduction in performing the single task, particularly when the tracking task is easy. Consequently, this model leads to the expectation of performance impairment on the single task following transition from the dual-task and that such impairment will be maximal when the tracking task is easy. Data collection on this project will is in progress. In addition to performance measures, The NASA-TLX and DSSQ scales is being used to measure workload and stress, respectively, and the Coping Inventory for Task Situations (CITS, Matthews & Campbell, 1998) is being employed to assess coping strategies. In addition, the TCD procedure is being utilized to provide a metabolic index of resource utilization. This study will serve as a Master's thesis for a second-year student, *Mr. Nathaniel Unger.*

3. To date, all of our work on cerebral hemodynamics has focused upon visual vigilance tasks. However, vigilance tasks can also be performed in the auditory modality and the sensory modality of signals is not a matter of indifference where vigilance is concerned. In addition to different transduction properties between the modalities, there are performance differences between them as well. Signal detection is greater in auditory than in visual tasks, inter-task correlations in performance efficiency are low, and task-induced stress is greater when observers must monitor visual than auditory displays (Warm & Jerison, 1984). Consequently, in order to establish the generality of our cerebral hemodynamic findings it is necessary to determine if changes in blood flow and cerebral oxygen saturation are similar in analogous auditory and visual vigilance tasks. That determination is one goal for this study. In addition, the study will also explore the effects of background event asymmetry, another powerful psychophysical parameter in vigilance, on cerebral blood flow and oxygen saturation. Instrumentation for the study is currently being developed. The study will serve as a Master's thesis for a first-year student, *Mr. Tyler Shaw.*

4. A key finding in the area of visual search is the phenomenon of search asymmetry, wherein it is easier to search for the presence of a target than for its absence (Palmer, 2002). Recent studies by Mark Scerbo at Old Dominion University have extended this effect to vigilance performance and demonstrated that the overall level of signal detection is greater and the vigilance decrement less severe when observers must detect targets defined by the presence as compared to the absence of critical features. Since signal difficulty effects in vigilance are exacerbated by increments in event rate, the present study examines the effects of increments in event rate on the search asymmetry effect in vigilance. In addition, since the asymmetry effect is explained by the argument that searching for feature absence is more capacity demanding than searching for feature presence (Palmer, 2002), this study is also designed to examine blood flow differences in the feature present/absent conditions using the TCD procedure as well as task-induced workload in these conditions using the NASA Task-Load Index and the Dundee multidimensional stress state questionnaire. Data collection for the study is currently in progress. The study will serve as a doctoral dissertation for a fourth-year student, *Mr. Todd Hollander*.

STANDARD

NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION

AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

AGREEMENT TITLE: Operator Under Stress Research

AGREEMENT NUMBER: NCRADA-**NAVAIR ORLANDO TSD-02-037**

AGREEMENT ADMINISTRATORS:

NAVAIR ORLANDO TSD

Technology Transfer Office:	**Ms. Patricia R. Nimmo, AIR 49T, (407) 380-8225**
Legal Counsel:	**Robert W. Adams, AIR 772P, (407) 380-8211**
Principal Investigator:	**Ron Wolff, AIR 4962, (407) 380-4583**

UCF MURI / OPUS

Preferred Contact:	**Mrs. Kim Smith, UCF Office of Research, (407) 823-3062**
Principal Investigator:	**Dr. Peter Hancock, Director, UCF MURI/OPUS Laboratory, (407) 823-2310**

This Page Intentionally Left Blank

TABLE OF CONTENTS

PREAMBLE

Article 1.
DEFINITIONS 2

1.1 "Agreement" 2
1.2 "Classified Information" 2
1.3 "Collaborator" 2
1.4 "Controlled Unclassified Information (CUI)" 2
1.5 "Cooperative Work" 2
1.6 "Data" 2
1.7 "Effective Date" 2
1.8 "Exclusive License" 2
1.9 "Government" 2
1.10 "Government Purpose Rights" 2
1.11 "Information" 2
1.12 "Intellectual Property" 2
1.13 "Invention" 2
1.14 "Invention Disclosure" 3
1.15 "Made" 3
1.16 "Militarily Critical Technologies (MCT)" 3
1.17 "Non-Subject Data" 3
1.18 "Non-Subject Invention" 3
1.19 "Patent Application" 3
1.20 "Principal Investigator (PI)" 3
1.21 "Proprietary Information" 3
1.22 "Restricted Access Information" 3
1.23 "Subject Data" 3
1.24 "Subject Invention" 3
1.25 "Tangible Property" 4
1.26 "Unlimited Rights" 4

Article 2. OBJECTIVES 4

Article 3. RESPONSIBILITIES 4

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities 4
3.2 **UCF MURI / OPUS Laboratory** Personnel and facilities 5
3.3 Security Regulations and Directives 5

Article 4. REPRESENTATIONS AND WARRANTIES 5

4.1 **NAVAIR ORLANDO TSD** Representations and Warranties 5
4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Representations and Warranties 5
4.3 Joint Representations 6

Article 5. FUNDING 7

5.1 Payment Schedule..... 7
5.2 Insufficient and Excess Funds..... 7
5.3 No New Commitments..... 8
5.4 Accounting Records..... 8

Article 6. REPORTS AND PUBLICATIONS..... 8

6.1 Final Reports..... 8
6.2 Agreement to Confer Prior to Publication or Public Disclosure..... 9
6.3 Classified Information 9

Article 7. INTELLECTUAL PROPERTY..... 9

7.1 Data..... 9
7.1.1 General Provisions Applying to All Data..... 9
7.1.1.1 Ownership..... 9
7.1.1.2 No Implied License..... 9
7.1.1.3 Marking of Data..... 10
7.1.1.3.1 Data Provided With Less Than Unlimited Rights.. 10
7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information..... 10
7.1.1.3.3 Data That are Subject to 35 USC 205..... 11
7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted..... 11
7.1.1.4 Protection of Data..... 11
7.1.1.5 Release of Data Under the Freedom of Information Act..... 11
7.1.2 Subject Data..... 12
7.1.2.1 Delivery of Requested Subject Data..... 12
7.1.2.2 Rights in Subject Data..... 12
7.1.3 Rights in Non-Subject Data..... 12
7.2 Copyrights..... 13
7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory**..... 13
7.2.2 Copyright License to the Government..... 13
7.2.3 Copyright Statement..... 13
7.3 Trademarks and Service Marks..... 13
7.3.1 Ownership of Trademarks and Service Marks..... 13
7.3.2 Obligation of Employees to Report Trademarks and Service Marks.... 13
7.3.3 Obligation of Collaborators to Notify Each Other..... 13
7.3.4 Responsibility for Filing an Application for Trademark or Service Mark..... 14
7.3.5 License to Use Trademark or Service Mark..... 14
7.4 Subject Inventions..... 14
7.4.1 Obligation to Report Subject Inventions..... 14
7.4.1.1 Collaborators' Instruction to Employees..... 14
7.4.1.2 Timely Invention Disclosure by Inventors..... 14
7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator..... 14
7.4.2 Determination of Subject Inventions..... 14
7.4.3 Title to and Ownership of Subject Inventions..... 14

7.4.4 Filing of Patent Applications..... 15
7.4.4.1 Filing of Patent Applications on Solely Made Inventions..... 15
7.4.4.2 Filing of Patent Applications on Jointly Made Inventions..... 15
7.4.4.3 Preserving Intellectual Property Rights..... 15
7.4.4.4 Filing Deadlines..... 15
7.4.4.5 Copies and Inspection..... 15
7.4.4.5.1 Copies of Prosecution Papers..... 15
7.4.4.5.2 Access to Patent Application File and Right to Make Copies..... 16
7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application..... 16
7.4.5 Nonexclusive License to Subject Inventions..... 16
7.4.5.1 Nonexclusive License Grant..... 16
7.4.5.2 Confirmatory Nonexclusive License Agreement..... 16
7.4.6 Option for Exclusive License to Subject Inventions..... 16
7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement. 17
7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions..... 17
7.4.8.1 Exclusive Licenses and Exclusive License Option..... 17
7.4.8.2 Nonexclusive Licenses..... 17
7.5 Non-Subject Inventions..... 18
7.5.1 Ownership of Non-Subject Inventions..... 18
7.5.2 Rights Under Other Agreements..... 18
7.5.3 No License to Non-Subject Inventions..... 18
7.6 Research License..... 18

Article 8. TANGIBLE PROPERTY..... 18

8.1 Title to Preexisting Tangible Property..... 18
8.2 Tangible Property Purchased by Collaborators to Perform Cooperative Work... 18
8.3 Title to Developed Tangible Property..... 19
8.4 Tangible Property Operational and Disposition Costs..... 19
8.5 Disposal of Tangible Property..... 19

Article 9. LIABILITY..... 19

9.1 Extent of Government Liability..... 19
9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Liability..... 19
9.3 *Force Majeure*..... 20

Article 10. GENERAL PROVISIONS..... 20

10.1 Characteristics of the Agreement..... 20
10.1.1 Entire Agreement..... 20
10.1.2 Severability..... 20
10.1.3 Interpretation of Headings..... 20
10.2 Agreement Between Collaborators..... 20
10.2.1 Governing Laws..... 20
10.2.2 Independent Parties/Entities..... 20
10.2.3 Assignment /Subcontracting..... 20

10.2.4 Disputes..... 21
10.2.4.1Settlement and Resolution..... 21
10.2.4.2Continuation of Cooperative Work..... 21
10.2.5 Waivers..... 21
10.2.6 Use of Name or Endorsements..... 22
10.3 Environment, Safety, and Health..... 22
10.4 U.S. Competitiveness..... 22
10.5 Public Release of This Agreement..... 22

Article 11. MODIFICATIONS AND NOTICES..... 22

11.1 Amendments..... 22
11.2 Termination..... 23
11.2.1 Termination by Mutual Consent..... 23
11.2.2 Unilateral Termination..... 23
11.3 Notices..... 23

Article 12. SURVIVING PROVISIONS..... 24

Article 13. DURATION..... 24

Article 14. SIGNATURES..... 24

Appendix A
Statement of Work

Appendix B
Confirmatory License Agreement

STANDARD
NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

PREAMBLE

Under authority of the U.S. Federal Technology Transfer Act of 1986 (Public Law 99-502, 20 October 1986, as amended), **NAVAIR ORLANDO TSD (Naval Air Warfare Center Training Systems Division)**, located at **12350 Research Parkway, Orlando, Florida 32826-3275**, and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, whose administrative offices are located at **12443 Research Parkway, Suite 207, Orlando, FL 32826-3252**, enter into this Cooperative Research and Development Agreement (CRADA), which shall be binding upon the Collaborators and their assignees according to the clauses and conditions hereof and for the term and duration set forth.

The U.S. Federal Technology Transfer Act of 1986, as amended, provides for making the expertise, capabilities, and technologies of U.S. Federal laboratories accessible to other Federal agencies; units of State or local government; industrial organizations (including corporations, partnerships and limited partnerships, and industrial development organizations); public and private foundations; nonprofit organizations (including universities); or other persons in order to improve the economic, environmental, and social well-being of the United States by stimulating utilization of U.S. Federally funded technology developments and/or capabilities.

NAVAIR ORLANDO TSD has extensive expertise, capabilities, and information in **Virtual Environments for Operator Training**, and, in accordance with the U.S. Federal Technology Transfer Act, desires to make this expertise and technology available for use in the public and private sectors.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory has the interest, resources, capabilities, and technical expertise to transition the results of Naval research and development for public use.

NOW THEREFORE, the Collaborators agree as follows.

Article 1. DEFINITIONS

As used in this Agreement, the following terms shall have the meanings defined below, which are equally applicable to both the singular and plural forms of nouns or any tense of verbs.

1.1 "Agreement" means this Cooperative Research and Development Agreement (CRADA) with its Appendices.

1.2 "Classified Information" means all Data classified in accordance with the national security laws of the United States.

1.3 "Collaborator" means the Navy participant or the Non-Navy participant represented and bound by the signatories of this Agreement.

1.4 "Controlled Unclassified Information (CUI)" means Government Data, Information, or materials provided to or resulting from this Agreement that may be export controlled, sensitive, for official use only, or otherwise protected by law, executive order, or regulation.

1.5 "Cooperative Work" means research, development, engineering, or other tasks performed under this Agreement by **NAVAIR ORLANDO TSD** or **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** working individually or together, pursuant to the Objectives (Article 2) and the Statement of Work (Appendix A).

1.6 "Data" means recorded information of any kind regardless of the form or method of the recording, including computer software.

1.7 "Effective Date" means the date of the last signature of the Collaborators executing this Agreement.

1.8 "Exclusive License" means the grant by the owner of Intellectual Property of the exclusive right to make, use, or sell a patented invention.

1.9 "Government" means the Government of the United States of America.

1.10 "Government Purpose Rights" means the right of the Government to use, duplicate, or disclose Data, in whole or in part, and in any manner, for Government purposes only, and to have or permit others to do so for Government purposes only. Government Purpose Rights includes competitive procurement, but does not include the right to have or permit others to use Data for commercial purposes.

1.11 "Information" means all data, trade secrets, and commercial and financial information. (Chapter 5 Subsection II of Title 5 USC)

1.12 "Intellectual Property" means the property of ideas, examples of which include, but are not limited to, patents, trademarks, copyrights, and trade secrets.

1.13 "Invention" means any invention or discovery that is or may be patentable or otherwise protected under Title 35, United States Code, or any novel variety of plant that is or may be patentable under the Plant Variety Protection Act. (15 USC 3703(9))

1.14 "Invention Disclosure" means the document identifying and describing to organizational management the Making of an Invention.

1.15 "Made" when used in conjunction with any Invention means the conception or first actual reduction to practice of such Invention. (15 USC 3703(10))

1.16 "Militarily Critical Technologies (MCT)" means those technologies identified in the Militarily Critical Technologies List and under the Export Administration Act of 1979, as amended.

1.17 "Non-Subject Data" means any Data that are not Subject Data.

1.18 "Non-Subject Invention" means any Invention that is not a Subject Invention.

1.19 "Patent Application" means an application for patent protection for an Invention with any domestic or foreign patent-issuing authority.

1.20 "Principal Investigator (PI)" means that person having the responsibility for the performance of the Cooperative Work on behalf of a Collaborator.

1.21 "Proprietary Information" means information that embodies trade secrets developed at private expense or business, commercial, or financial information that is privileged or confidential provided that such information:

is not known or available from other sources without obligations concerning its confidentiality;

has not been made available by the owners to others without obligation concerning its confidentiality;

is not already available to the Government without obligation concerning its confidentiality; and

has not been developed independently by persons who have had no access to the information. (FAR/DFARS Definition)

1.22 "Restricted Access Information" means Subject Data generated by **NAVAIR ORLANDO TSD** that would be Proprietary Information if the Information had been obtained from a non-Federal Collaborator participating in a CRADA (15 USC 3710a). Under 15 USC 3710a(c)(7)(B), the Collaborators mutually may agree to provide appropriate protection to Subject Data generated by **NAVAIR ORLANDO TSD** (Restricted Access Information) against public dissemination or release under the Freedom of Information Act (FOIA) for a period of up to five (5) years after development of the Information.

1.23 "Subject Data" means that Data first recorded in the performance of the Cooperative Work.

1.24 "Subject Invention" means any Invention Made in the performance of the Cooperative Work.

1.25 "Tangible Property" means personal or real property that can be physically touched or held.

1.26 "Unlimited Rights" means the right to use, modify, reproduce, release, disclose, perform, or display Data or Computer Programs in whole or in part, in any manner and for any purpose whatsoever, and to have or permit others to do so.

Article 2. OBJECTIVES

Background

NAVAIR ORLANDO TSD has been conducting Navy sponsored research in virtual environments for operator/small arms training for over 15 years. Results of this research have already been transitioned for many military training applications, and have also been the basis for technology transition in use by law enforcement. The Small Arms Simulation Training (SAST) laboratory at **NAVAIR ORLANDO TSD** now has in use an extensive range of software and data bases that have resulted from in-house **NAVAIR ORLANDO TSD** research work.

The University of Central Florida (UCF), with plans to establish a new laboratory for Human Factors research related to Homeland Defense, became aware of the work already done in **NAVAIR ORLANDO TSD's** SAST laboratory. Discussions were immediately begun to explore the mutual benefits of an agreement to work collaboratively.

The primary objective of this work will be to enhance mutual understand of operator performance under stress and develop possible applications of human factors research in this area for military training, and for homeland defense and related technologies

Benefit to Navy Partner: Research using **NAVAIR ORLANDO TSD** developed or enhanced source code may be expected to result in further enhancements to software and databases, at no expense to the government, and with direct use in military simulation-based training. Additionally, the anticipated UCF human factors research projects may be expected to have direct application and benefit for military training. Further, expertise will be provided to UCF on a cost basis, leveraging Navy research dollars.

Benefit to Non-Navy partner: **NAVAIR ORLANDO TSD** developed or enhanced source code for scenario generation, rural and urban databases, weapon tracking, image generation, networking/communication, etc., can provide the kind of virtual environment test bed required for planned human factors research, and can greatly accelerate work toward research objectives.

Article 3. RESPONSIBILITIES

The Collaborators shall provide personnel, facilities, and equipment necessary for, and shall perform, the Cooperative Work.

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities

The Cooperative Work done by **NAVAIR ORLANDO TSD** will be performed under the program guidance of **Ron Wolff, PI, NAVAIR ORLANDO TSD, AIR 4962**, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed

within the facilities of **NAVAIR ORLANDO TSD** or done on behalf of **NAVAIR ORLANDO TSD** by third parties in support of this Agreement.

UCF MURI / OPUS Laboratory personnel who perform Cooperative Work at **NAVAIR ORLANDO TSD** facilities will be supervised by the **UCF MURI / OPUS Laboratory** PI.

3.2 **UCF MURI / OPUS Laboratory** Personnel and Facilities

The Cooperative Work done by the **UCF MURI / OPUS Laboratory** will be performed under the program guidance of **Dr. Peter Hancock, UCF MURI / OPUS Laboratory** PI, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed within the facilities of the **UCF MURI / OPUS Laboratory** or done on behalf of the **UCF MURI / OPUS Laboratory** by third parties in support of this Agreement.

NAVAIR ORLANDO TSD personnel who perform Cooperative Work at the **UCF MURI / OPUS Laboratory** facilities will be supervised by **NAVAIR ORLANDO TSD** PI.

3.3 Security Regulations and Directives

Each Collaborator will abide by the safety and security regulations and directives of the host facility in which the Cooperative Work is being performed.

Article 4. REPRESENTATIONS AND WARRANTIES

4.1 **NAVAIR ORLANDO TSD**'s Representations and Warranties

NAVAIR ORLANDO TSD hereby warrants and represents to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** as follows:

4.1.1 **NAVAIR ORLANDO TSD** is a Federal laboratory of the U.S. Department of the Navy (Navy) as defined by 15 USC 3710a(d)(2)(A) and Department of Defense Instruction 5535.8, dated May 14, 1999.

4.1.2 The performance of the activities specified by this Agreement is consistent with the **Virtual Environments for Operator/Small Arms Training** and technology transfer missions of **NAVAIR ORLANDO TSD** (15 USC 3710a).

4.1.3 The Department of the Navy official executing this Agreement for **NAVAIR ORLANDO TSD** has the requisite power and authority to enter into this Agreement and to bind **NAVAIR ORLANDO TSD** to perform according to the terms of this Agreement.

4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Representations and Warranties

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory hereby warrants and represents to **NAVAIR ORLANDO TSD** as follows:

4.2.1 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not directly or indirectly controlled by a foreign company or government (Executive Order 12591, Section 4 (a)). **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, as of the Effective Date of this Agreement, is a part of a **university** duly organized, validly existing, and in good standing under the laws of **the State of Florida.**

4.2.2 The official executing this Agreement for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** has the requisite power and authority to enter into this Agreement and to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to perform according to the terms of this Agreement.

4.2.3 The University of Central Florida and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** have taken all actions required by law, or otherwise, to authorize the execution and delivery of agreements, such as this Agreement.

4.2.4 The execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any agreement binding on **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. Furthermore, the execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any valid order of any court, or any regulatory agency or other body having authority to which **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is subject.

4.2.5 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not presently subject to debarment or suspension by any agency of the Government. Should **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** be debarred or suspended during the term of this Agreement or thereafter, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** will notify **NAVAIR ORLANDO TSD** within thirty (30) days of receipt of a final notice. **NAVAIR ORLANDO TSD** may then elect to terminate this Agreement and any licenses and options granted under this Agreement.

4.2.6 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is not** a small business as defined in 15 USC 632 and implementing regulations (13 CFR 121.101 et seq.) of the Administrator of the Small Business Administration.

4.3 Joint Representations

The Collaborators make the following representations.

4.3.1 There is no express or implied warranty as to any research, Invention, or product, whether tangible or intangible. In particular, the Collaborators make no express or implied warranty as to the merchantability or fitness for a particular purpose of any research, Invention, or product, whether tangible or intangible. Likewise, the Collaborators make no express or implied warranty as to any Cooperative Work, Subject Invention, Subject Data, or other product resulting from the Cooperative Work.

4.3.2 The use and dissemination of Information and materials exchanged under this Agreement will be in accordance with all U.S. laws and regulations, including those pertaining to national security and export control. Nothing in this Agreement shall be construed as a license to export Information or to permit any disclosure in violation of law, regulation, or Department of Defense policy. The exporting Collaborator is responsible for obtaining any export licenses that may be required by U.S. Federal law.

Article 5. FUNDING

5.1 Payment Schedule

Consistent with the information in Appendix A, **NAVAIR ORLANDO TSD** will provide source code/databases developed or enhanced by **NAVAIR ORLANDO TSD** to the **MURI / OPUS** lab at no cost to **UCF**. If, at a later date during the period of this CRADA, there is a mutually determined need for additional expertise from **NAVAIR ORLANDO TSD**, then the following will apply. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees to pay **NAVAIR ORLANDO TSD** on a cost basis, as mutually determined to be required, during the period of this CRADA, as specified in Appendix A.

Checks will be payable to:

NAVAIR ORLANDO TSD, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments

Each check and its cover correspondence shall refer to Navy CRADA number "NCRADA-**NAVAIR ORLANDO TSD-02-037**."

Checks will be mailed to:

NAVAIR ORLANDO TSD, 12350 Research Parkway (AIR 4.9T Nimmo), Orlando, Florida 32826-3275, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments.

5.2 Insufficient and Excess Funds

NAVAIR ORLANDO TSD may discontinue performance under this Agreement if the funds provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** for performance by **NAVAIR ORLANDO TSD** are insufficient or are not provided as specified in Article 5.1. In the event **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to tender the Government the required payment within fifteen (15) days after its respective due date, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall be in default under this Agreement for failure to make payments. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on**

behalf of the MURI / OPUS Laboratory is in default for this reason, NAVAIR ORLANDO TSD shall notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** does not cure the default within fifteen (15) days of mailing date of notice, **NAVAIR ORLANDO TSD** may proceed to terminate the Agreement in accordance with Article 11.2.2, may cancel any option for an Exclusive License to a Subject Invention, and may terminate any Exclusive License granted pursuant to this Agreement.

Funds that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** paid under Article 5.1 and that **NAVAIR ORLANDO TSD** has not obligated or expended at the time of completion, expiration, or termination of this Agreement shall be returned to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** after **NAVAIR ORLANDO TSD's** submission of a final fiscal report to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**.

5.3 No New Commitments

NAVAIR ORLANDO TSD shall make no new commitments concerning this Agreement after receipt of a written termination notice from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in accordance with Article 11.2 and shall, to the extent practicable, cancel all outstanding commitments by the termination date. Should such cancellation result in any costs incurred by **NAVAIR ORLANDO TSD, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that such costs shall be chargeable against any funding that it provided to **NAVAIR ORLANDO TSD.**

5.4 Accounting Records

NAVAIR ORLANDO TSD shall maintain current accounts, records, and other evidence supporting all its expenditures against funding provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement and shall retain such records for at least twelve (12) months after the completion, expiration, or termination of this Agreement. **NAVAIR ORLANDO TSD** shall provide **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** a financial report within four (4) months after completion, expiration, or termination of this Agreement.

Article 6. REPORTS AND PUBLICATIONS

6.1 Final Report

The Collaborators shall submit to each other a final report within four (4) months of the completion, termination, or expiration of this Agreement that includes the results obtained and a list of all Subject Inventions Made.

6.2 Agreement to Confer Prior to Publication or Public Disclosure

The Collaborators agree to confer and consult prior to any publication or public disclosure of Subject Data to ensure that no Proprietary Information, Restricted Access Information, Government Classified Information, CUI, or MCT Information is released and that patent rights are not compromised. Prior to any such publication or public disclosure of Subject Data, each Collaborator shall be offered a period not less than fifteen (15) days and not to exceed thirty (30) days, unless otherwise mutually agreed in writing by the Collaborators, to review any proposed abstract, publication, presentation, or other document for public disclosure that contains Subject Data. For the purposes of this Article, the term "disclosure" shall include, but not be limited to, submission of any manuscript for peer review prior to publication. It is the responsibility of the Collaborator intending to make public disclosure of Subject Data to notify the other Collaborator of such intent.

If a Collaborator objects to a proposed public disclosure, that Collaborator must so notify the other Collaborator within thirty (30) days of the date of notice of intent to disclose publicly. If no objection is received by the Collaborator intending to make public disclosure, concurrence is assumed. If a Collaborator objects on the grounds that patent rights may be compromised, a Patent Application must be filed by the responsible Collaborator within ninety (90) days of the date of notification of intent to make public disclosure, or by another date mutually agreed to by the Collaborators. If a Collaborator objects to the release of Information on the grounds that the Information is Proprietary Information, Restricted Access Information, or Information whose dissemination is restricted by U.S. security laws or regulations, the disclosure shall be postponed until the Information no longer meets the definitions of Proprietary Information, Restricted Access Information, or is no longer covered by U.S. security laws or regulations.

6.3 Classified Information

Any presentation that includes Subject Data that are Classified Information or otherwise restricted Data must have prior review and approval by **NAVAIR ORLANDO TSD** pursuant to the pertinent security laws, regulations, and directives.

Article 7. INTELLECTUAL PROPERTY

7.1 Data

7.1.1 General Provisions Applying to All Data

7.1.1.1 Ownership

Each Collaborator shall have title to all Data generated by that Collaborator.

7.1.1.2 No Implied License

Unless otherwise specifically provided, the Collaborators agree that the exchange of Data of any kind does not confer a license to any Invention claimed in any

patent or Patent Application or to the subject matter of any copyright, trademark/service mark, or other form of Intellectual Property protection.

7.1.1.3 Marking of Data

7.1.1.3.1 Data Provided With Less Than Unlimited Rights

Each Collaborator shall mark all Data that it provides with less than Unlimited Rights with a marking that clearly identifies the limited rights.

7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall place a proper proprietary marking on each medium used for recording Data that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** delivers to **NAVAIR ORLANDO TSD** under this Agreement that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** asserts is Proprietary Information. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall request in writing if it wishes Subject Data generated by **NAVAIR ORLANDO TSD** to be marked as Restricted Access Information. The Collaborators together shall confer to determine if such marking is appropriate, with reference to the Definitions of Article 1. If the Collaborators mutually agree to the marking then:

(a) For Non-Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory – NAVAIR ORLANDO TSD** MAY USE ONLY FOR PURPOSE OF CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(b) For Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** – GOVERNMENT HAS GOVERNMENT PURPOSE RIGHTS UNDER CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(c) For Data that are Restricted Access Information, the marking shall read:

"RESTRICTED ACCESS INFORMATION – PROTECT IN ACCORDANCE WITH CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**" UNTIL **SEPTEMBER 2005**".

7.1.1.3.3 Data That are Subject to 35 USC 205

NAVAIR ORLANDO TSD shall mark Data it provides under this Agreement that disclose one or more Inventions in which the Government owns or may own a right, title or interest, and that are subject to confidentiality under 35 USC 205. Such Data shall be marked:

"**NAVAIR ORLANDO TSD** DATA PROTECTED FROM RELEASE OR DISCLOSURE UNDER 35 USC 205."

7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted

Each Collaborator shall mark all Data that are Classified Information, CUI, MCT, or otherwise restricted by U.S. security or export control laws or regulations that it provides under this Agreement.

7.1.1.4 Protection of Data

Except for the rights granted in Article 7.1.2.2, Data shall be protected in accordance with the proper markings of its owner and as provided by, at a minimum, the requirements of 15 USC 3710a. Proprietary Information will be protected only if it is properly marked as such. Information provided in intangible form that is Proprietary Information must be designated Proprietary Information at the time it is delivered, followed within fifteen (15) days by a writing summarizing the exact Information to be protected. The Collaborator receiving Information in an intangible form that is designated as Proprietary Information shall be responsible for protecting the Information as Proprietary Information during the fifteen (15) day notification period. After the fifteen (15) day period, if no written summary has been received, the receiving Collaborator need not continue to protect the Information received in intangible form.

Restricted Access Information shall be protected from public dissemination for up to five (5) years, as mutually agreed.

Classified Information, CUI, MCT, or otherwise restricted Information shall be protected in accordance with the security laws of the United States.

7.1.1.5 Release of Data Under the Freedom of Information Act

Data in the possession of **NAVAIR ORLANDO TSD** that are not marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or Restricted Access Information must be released by **NAVAIR ORLANDO TSD** where such release is required pursuant to a request under the Freedom of Information Act (FOIA) (5 USC 552). **NAVAIR ORLANDO TSD** shall protect Data that are properly marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI**

/ **OPUS Laboratory** or Restricted Access Information from release under the FOIA for as long as the marked Data meet the definition of CUI, Proprietary Information or Restricted Access Information. Prior to release of any such Data, **NAVAIR ORLANDO TSD** shall promptly notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** of any request for Data of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** regardless of whether the requested Data are marked Proprietary Information or as required under the public records law.

7.1.2 Subject Data

7.1.2.1 Delivery of Requested Subject Data

Each Collaborator shall have the right to review and receive delivery of all Subject Data generated by the other Collaborator. Requested Subject Data shall be delivered to the requesting Collaborator within fifteen (15) days of the request.

7.1.2.2 Rights in Subject Data

Except as represented in Article 4.3.2, the Collaborators shall have Unlimited Rights in all Subject Data that are not Proprietary Information or Restricted Access Information. Notwithstanding 15 USC 3710a, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** grants Government Purpose Rights in any Subject Data furnished by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to **NAVAIR ORLANDO TSD** under this Agreement that are properly marked as Proprietary Information. The Government has Government Purpose Rights in Subject Data that are Restricted Access Information.

7.1.3 Rights in Non-Subject Data

The Collaborators shall have Unlimited Rights in any Non-Subject Data delivered under this Agreement that are not Proprietary Information.

NAVAIR ORLANDO TSD has a limited right to use, reproduce, and disclose only to Government employees for use in support of the Cooperative Work any Non-Subject Data that are properly marked as Proprietary Information and are provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement. Such Proprietary Information can be used only for the purpose of performing the Cooperative Work unless consent to other use or disclosure is obtained from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in writing.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall have a limited right to use, reproduce, or disclose Non-Subject Data that may describe one or more Inventions in which the Government owns or may own a right, title or interest, if such Non-Subject Data are provided by **NAVAIR ORLANDO TSD** under this Agreement. In accordance with 35 USC 205, such Non-Subject Data are to be held in confidence. Such Non-Subject Data shall be properly marked by **NAVAIR ORLANDO TSD** and the limited rights of **The UCF Office of Research, UCF on behalf of its**

Board of Trustees on behalf of the MURI / OPUS Laboratory shall be defined by a separate non-disclosure agreement.

7.2 Copyrights

7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory may copyright works of authorship prepared pursuant to this Agreement if eligible for copyright protection under Title 17 USC.

7.2.2 Copyright License to the Government

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory grants to the Government a nonexclusive, irrevocable, paid-up license in copyrighted works of authorship, including software (17 USC 106) prepared pursuant to this Agreement for any purpose, consistent with the rights in Data described in Article 7.1.

7.2.3 Copyright Statement

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall include the following statement on any text, drawing, mask work or other work of authorship, that may be copyrighted under 17 USC, that is created in the performance of this Agreement:

"The U.S. Government has a copyright license in this work pursuant to a Cooperative Research and Development Agreement with **NAVAIR ORLANDO TSD.**"

7.3 Trademarks and Service Marks

7.3.1 Ownership of Trademarks and Service Marks

The Collaborator first establishing a trademark or service mark for goods or services with which the mark is used shall be considered the owner of the mark.

7.3.2 Obligation of Employees to Report Trademarks and Service Marks

Employees of both Collaborators shall report the adoption of a trademark or service mark associated with the Cooperative Work to their employer within thirty (30) days of the first use of the mark. Use includes internal use of any product or service of the Cooperative Work.

7.3.3 Obligation of Collaborators to Notify Each Other

Each Collaborator shall notify the other Collaborator within thirty (30) days of their employee's report of the first use of a trademark or service mark.

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

Each Collaborator shall be entitled to own the Subject Inventions of its employees. Each Collaborator shall cooperate with the other Collaborator to obtain inventor signatures on Patent Applications, assignments or other documents required to secure Intellectual Property protection. For any Invention Made jointly by employees of the Collaborators, each Collaborator shall have ownership of the Subject Invention in the form of an undivided interest.

7.4.4 Filing of Patent Applications

7.4.4.1 Filing of Patent Applications on Solely Made Inventions

Each Collaborator has primary responsibility for filing Patent Applications on the Subject Inventions of its employee(s).

Notwithstanding such primary responsibility, by mutual agreement, the Collaborators may identify which Collaborator shall file a Patent Application on any Subject Invention.

7.4.4.2 Filing of Patent Applications on Jointly Made Inventions

In the case of an Invention jointly Made by employees of both Collaborators, the Collaborators shall confer and agree as to which Collaborator will file any Patent Application. Officers of the non-filing Collaborator shall cooperate with the filing Collaborator to obtain signatures on documents that are needed to file a Patent Application.

7.4.4.3 Preserving Intellectual Property Rights

The Collaborator responsible for filing of a Patent Application on any Subject Invention shall file such Patent Application at least sixty (60) days prior to any bar date or one year from the date the Invention Disclosure was received, whichever comes first. If no Patent Application is filed within the specified time period, the other Collaborator may assume control of filing the Patent Application and take title to the Subject Invention on ten (10) days written notification. The Collaborator that relinquished the responsibility to file shall retain a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Subject Invention practiced throughout the world by or on its behalf.

7.4.4.4. Filing Deadlines

The Collaborator responsible for filing any Patent Application for a Subject Invention shall notify the other Collaborator of all filing deadlines for prosecution of any Patent Application and maintenance of any patents on the Subject Invention. Notwithstanding the primary responsibility defined in Article 7.4.4.1, sixty (60) days prior to any filing deadline, the Collaborators shall confer to determine if the filing Collaborator intends to respond to the filing deadline. The non-filing Collaborator will be permitted to take action if the filing Collaborator declines.

7.4.4.5 Copies and Inspection

7.4.4.5.1 Copies of Prosecution Papers

Each Collaborator filing a Patent Application on a Subject Invention shall provide the other Collaborator with a copy of any communication relating

to prosecution of said Patent Application within thirty (30) days of receipt of such communication.

7.4.4.5.2 Access to Patent Application File and Right to Make Copies

Upon written request, the filing Collaborator shall give the other Collaborator an Associate Power of Attorney, with authorization to access the Patent Application, make copies, and, in the event the filing Collaborator fails or declines to take action, do all that is necessary to secure Intellectual Property protection for the Subject Invention.

7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application

In the event both Collaborators decline to file a Patent Application on a Subject Invention, the Government will renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

In the event both Collaborators decline to file a Patent Application on a Subject Invention, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** may, at its sole discretion, renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

7.4.5 Nonexclusive License to Subject Inventions

7.4.5.1 Nonexclusive License Grant

Each Collaborator grants to the other Collaborator a nonexclusive, irrevocable, paid-up license to practice a Subject Invention Made by employees of the granting Collaborator or have the Subject Invention practiced throughout the world by or on behalf of the other Collaborator. No nonexclusive license granted under this Agreement shall permit licensee to grant sublicenses.

7.4.5.2 Confirmatory Nonexclusive License Agreement

Each Collaborator has the obligation to provide a Confirmatory License Agreement (Appendix B) to the other Collaborator for each nonexclusive license within ninety (90) days of the date of filing.

7.4.6 Option for Exclusive License to Subject Inventions

NAVAIR ORLANDO TSD gives **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** the option of acquiring an Exclusive License for the field of use **(virtual environments for operator/small arms training)** in the Government's rights in any Subject Invention Made in whole or in part by a **NAVAIR ORLANDO TSD** employee. The license shall be for reasonable consideration. In

order to exercise this option, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must notify **NAVAIR ORLANDO TSD** in writing within one hundred and eighty (180) days of the filing of a Patent Application. Unless another time period is mutually agreed upon between the Collaborators, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must execute an Exclusive License to the Subject Invention within one hundred and eighty (180) days of election to exercise the option, or the Invention shall be made available for licensing by the public in accordance with 37 CFR Part 404.

Any Exclusive License granted by the Government in a Subject Invention is subject to the statutorily required reservation by the Government of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the Government (15 USC 3710a).

7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement

No license granted under this Agreement shall be assigned, licensed or otherwise disposed of except to the successor in interest of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which such license pertains.

7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions

7.4.8.1 Exclusive Licenses and Exclusive License Option

NAVAIR ORLANDO TSD may terminate any Exclusive License or cancel any option for an Exclusive License to a Subject Invention granted under this Agreement in the event that:

(a) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is in default for failure to make payment as agreed in Article 5; or

(b) The Agreement is terminated unilaterally by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**; or

(c) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to perform according to the Statement of Work (Appendix A); or

(d) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** becomes a foreign owned, controlled, or influenced (FOCI) organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.4.8.2 Nonexclusive Licenses

NAVAIR ORLANDO TSD shall terminate any nonexclusive license to a Subject Invention granted under this Agreement if **The UCF Office of Research,**

UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory becomes a FOCI organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.5 Non-Subject Inventions

7.5.1 Ownership of Non-Subject Inventions

Each Collaborator owns its Non-Subject Inventions.

7.5.2 Rights Under Other Agreements

Nothing in this Agreement is intended to change the rights in Intellectual Property acquired by the Collaborators in any other contract or agreement between the **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** and the Government.

7.5.3 No License to Non-Subject Inventions

This Agreement does not grant any Collaborator a license, express or implied, to any Non-Subject Invention.

7.6 Research License

Each Collaborator shall allow the other Collaborator to practice any of its Non-Subject Inventions for the purpose of performing the Cooperative Work.

No license, express or implied, for commercial application(s) is granted to either Collaborator in Non-Subject Inventions by performing the Cooperative Work.

For commercial application(s) of Non-Subject Inventions, a license must be obtained from the owner.

Article 8. TANGIBLE PROPERTY

8.1 Title to Preexisting Tangible Property

Each Collaborator shall retain title to all Tangible Property to which it had title prior to the Effective Date of this Agreement.

8.2 Tangible Property Purchased by Collaborators to Perform the Cooperative Work

Each Collaborator shall retain title to all Tangible Property that it purchases during the period of this Agreement. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** cannot take title to any Government Tangible Property under this Agreement. Collaborator consumables to be used in the Cooperative Work of this Agreement are the property of the purchasing Collaborator until consumed.

8.3 Title to Developed Tangible Property

All Tangible Property developed under this Agreement with all components purchased by one Collaborator shall be the property of that Collaborator. Tangible Property having any component purchased by **NAVAIR ORLANDO TSD** shall be the property of the Government, unless such Tangible Property can reasonably be separated without damage to the other individual components. After this Agreement is completed, expired, or terminated, if separation of components can be made without damage, the Collaborators may, by mutual agreement, separate the Tangible Property into its components and the separated components shall remain the property of the Collaborator that purchased them.

8.4 Tangible Property Operational and Disposition Costs

During the period of and upon completion, expiration, or termination of this Agreement, each Collaborator shall be responsible for all costs of maintenance, removal, storage, repair, disposal, and shipping of all Tangible Property to which it has title.

8.5 Disposal of Tangible Property

Disposal of Tangible Property shall be in accordance with applicable U.S. Federal, State, and local property disposal laws, environmental laws, and regulations.

Article 9. LIABILITY

9.1 Extent of Government Liability

The Government shall be solely liable for the negligent or wrongful acts of its officers and employees to the extent provided for in the Federal Tort Claims Act (28 USC 2671 et. seq.) and in other applicable laws and regulations of the United States that specifically waive sovereign immunity. Nothing in this Agreement shall be construed as a waiver of the sovereign immunity of the United States.

9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Liability

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is solely responsible for its actions and the actions of those acting for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in the performance of this Agreement and for any damages that may arise from any suit, action, or claim, and for any costs from or incidental to any suit, action, or claim, including but not limited to settlement and defense costs. Further, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that in any suit, action or claim brought by anyone not a party to this Agreement based on actions of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not pursue any actions to enter the Government as a party in such suit, action or claim unless the Government has some liability under the Federal Tort Claims Act.

9.3 *Force Majeure*

No Collaborator shall be liable for the consequences of any *force majeure* that (1) is beyond its reasonable control; (2) is not caused by the fault or negligence of such Collaborator; (3) causes such Collaborator to be unable to perform its obligations under this Agreement; and (4) cannot be overcome by the exercise of due diligence. In the event of the occurrence of a *force majeure*, the Collaborator unable to perform shall promptly notify the other Collaborator. The Collaborators shall suspend performance only for such period of time as is necessary to overcome the result(s) of the *force majeure* and shall use their best efforts to resume performance as quickly as possible.

Article 10. GENERAL PROVISIONS

10.1 Characteristics of the Agreement

10.1.1 Entire Agreement

This Agreement constitutes the entire agreement between the Collaborators concerning the Cooperative Work and supersedes any prior understanding or written or oral agreement relative to the Cooperative Work.

10.1.2 Severability

The illegality or invalidity of any Article of this Agreement shall not impair, affect, or invalidate any other Article of this Agreement.

10.1.3 Interpretation of Headings

Headings of the Articles of this Agreement are for convenience of reference only and do not form a part of this Agreement and shall in no way affect the interpretation thereof.

10.2 Agreements Between Collaborators

10.2.1 Governing Laws

United States Federal Laws shall govern this Agreement for all purposes.

10.2.2 Independent Parties/Entities

The relationship of the Collaborators to this Agreement is that of independent parties and not as agents of each other, partners, or participants in a joint venture. Each Collaborator shall maintain sole and exclusive control over its personnel and operations.

10.2.3 Assignment/Subcontracting

10.2.3.1 Neither Collaborator may allow third parties to perform any part of the Cooperative Work under this Agreement without express written consent of the other

Collaborator. If consent is obtained, the Collaborator requesting such consent shall remain fully responsible for the portion of the Cooperative Work to be accomplished under a third-party agreement, and the third party is not a Collaborator of this Agreement. Any third-party agreement to perform a portion of the Cooperative Work shall contain terms consistent with this Agreement.

10.2.3.2 This Agreement shall not be assigned or otherwise transferred by either Collaborator without the prior written consent of the other Collaborator, except to the successor of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which this Agreement pertains.

10.2.3.3 If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee is a U.S. company, and becomes, during the term of this Agreement or thereafter, directly or indirectly owned, controlled, or influenced by a foreign company or government (FOCI), then **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee shall promptly notify **NAVAIR ORLANDO TSD** to that effect.

10.2.4 Disputes

10.2.4.1 Settlement and Resolution

NAVAIR ORLANDO TSD and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agree to use reasonable efforts to reach a fair settlement of any dispute. If such efforts are unsuccessful, remaining issues in dispute will be referred to the signatories or their successors for resolution. If a dispute continues, the remaining issues may be submitted to the Chief of Naval Research (CNR), or the CNR designee, for resolution. This Agreement does not prevent any Collaborator from pursuing disputes in a U.S. Federal court of competent jurisdiction. No Collaborator will pursue litigation in a U.S. Federal court until after the CNR, or the CNR designee, decides the dispute, or until sixty (60) days after the dispute was first submitted to the CNR, or the CNR designee, whichever comes first.

10.2.4.2 Continuation of Cooperative Work

If payments or installment payments are to be made as stated under Article 5, **NAVAIR ORLANDO TSD** will not start or continue cooperative work until payments or installment payments are received.

10.2.5 Waivers

None of the provisions of this Agreement shall be considered waived by either Collaborator unless such waiver is given in writing to the other Collaborator, signed by the executing official of this Agreement or the official's successor having the authority to bind the Collaborator making the waiver. The failure of either Collaborator to insist upon strict performance of any of the terms and conditions herein, or failure or delay to exercise any rights provided herein or by law shall not be deemed a waiver of any right of either Collaborator under this Agreement.

10.2.6 Use of Name or Endorsements

Except as provided for in Article 7.2.3, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not use the name of **NAVAIR ORLANDO TSD** or any other Government entity on any product or service that is directly or indirectly related to either this Agreement or any patent license or assignment associated with this Agreement without the prior approval of **NAVAIR ORLANDO TSD**. By entering into this Agreement, **NAVAIR ORLANDO TSD** does not directly or indirectly endorse any product or service provided, or to be provided, by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, its successors, assignees, or licensees. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not in any way imply that the Department of the Navy endorses any such product or service.

10.3 Environment, Safety, and Health

Each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste in its custody during the course of this Agreement. At the conclusion of this Agreement, each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste still in its possession. Each Collaborator shall obtain at its own expense all necessary permits and licenses as required by U.S. Federal, State, and local law and shall conduct such handling, control, and disposition in a lawful and environmentally responsible manner. Each Collaborator is responsible for all required environmental, safety, and health compliance, notice, and monitoring related to its facility in accordance with U.S. Federal, State, and local law and regulations. Collaborators shall abide by the environmental, safety, and health directives of the host facility in which the Cooperative Work is being performed, and any U.S. Federal, State, or local laws and regulations pertaining to environment, safety, and health that are applicable to the host facility.

10.4 U.S. Competitiveness

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory agrees that any product, process, or service using Intellectual Property arising from the performance of this Agreement shall be manufactured substantially in the United States.

10.5 Public Release of This Agreement

This Agreement, without funding information (Article 5) and Appendices, may be released to the public.

Article 11. MODIFICATIONS AND NOTICES

11.1 Amendments

If a Collaborator wishes to modify this Agreement, the Collaborators shall confer in good faith to determine the desirability of such modification. Such modification shall not be effective until a written amendment is signed by both executing officials of this Agreement or their successors.

11.2 Termination

11.2.1 Termination by Mutual Consent

The Collaborators may elect to terminate this Agreement at any time by mutual consent. Such termination shall not be effective until a written termination agreement is signed by both executing officials of this Agreement or their successors.

11.2.2 Unilateral Termination

A Collaborator may unilaterally terminate this entire Agreement at any time by giving the other Collaborator written notice signed by the executing official of this Agreement or his/her successor, not less than thirty (30) days prior to the desired termination date. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** unilaterally terminates this Agreement, any option for an Exclusive License to a Subject Invention and any Exclusive License to a Subject Invention granted by or pursuant to this Agreement shall simultaneously be terminated.

11.3 Notices

All notices pertaining to or required by Articles of this Agreement, except those pertaining solely to the prosecution of any patent, trademark, or service mark, shall be in writing and shall be signed by an authorized representative of the Technology Transfer Office for **NAVAIR ORLANDO TSD** or the preferred contact for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, and all such notices shall be delivered by hand, sent by courier with proper registration, or sent by certified mail, return receipt requested, with postage prepaid, addressed as follows:

If to **NAVAIR ORLANDO TSD**:

Ms. Patricia R. Nimmo
Technology Transfer Project Manager
NAVAIR ORLANDO TSD (AIR 49T)
12350 Research Parkway
Orlando, FL 32826-3275

If to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

Ms. Kim Smith
Senior Contract Specialist
Office of Research
University of Central Florida
Orlando Tech Center
12443 Research Parkway, Suite 207
Orlando, Fl 32826-3252

A Collaborator shall notify the other Collaborator of a change of address in the manner set forth above.

Notices pertaining solely to the prosecution of any patent, trademark, or service mark related to this Agreement shall be in writing and shall be signed by and sent to the Collaborator's legal counsel for Intellectual Property. Legal counsel for Intellectual Property for each Collaborator shall send a copy of any such notice to the Technology Transfer Office for he **NAVAIR ORLANDO TSD**. If either Collaborator fails to identify such counsel upon request, then such notices shall be sent to the points of contact specified above.

Article 12. SURVIVING PROVISIONS

The Articles covering Definitions, Representations and Warranties, Funding, Reports and Publications, Intellectual Property, Tangible Property, Liability, General Provisions, Modifications and Notices, and Surviving Provisions shall survive the completion, termination, or expiration of this Agreement.

Article 13. DURATION

This Agreement expires three years after its Effective Date, unless otherwise extended in writing according to the provisions of Article 11.

Article 14. SIGNATURES

For **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

I, the undersigned, am duly authorized to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002

By: M.J. Soileau

Title: Vice President, Office of Research

For the Department of the Navy:

I, the undersigned, by 15 USC 3710a and Navy regulations, am duly authorized to bind the U.S. Navy to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002.

By:

R. A. MOHLER

Title: Captain, U.S. Navy
Commanding Officer

Navy Organization: NAVAIR ORLANDO, TRAINING SYSTEMS DIVISION

This Page Intentionally Left Blank

APPENDIX A

STATEMENT OF WORK
BETWEEN
NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND
THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

NAVAIR ORLANDO TSD will:

1. Provide, for use in the UCF MURI/OPUS laboratory, source code/databases developed or enhanced by NAVAIR ORLANDO TSD as part of Navy sponsored research and development, e.g., Host Software, Image Generator Software, Weapon Tracking Software, Networking/Communications Software, Data Collection/Analysis Software, Scenario Builder Software, Rural and Urban Databases

2. Provide expertise, on a cost basis and as mutually determined to be required, on utilization of government provided software and databases for UCF MURI/OPUS Laboratory research

3. Provide, on a cost basis and as mutually determined to be required, expertise for further hardware/software/database development, testing, and training applications

4. Participate, on a cost basis, in mutually selected specific research projects conducted under the auspices of the UCF MURI/OPUS Laboratory

5. Provide access to NAVAIR ORLANDO TSD customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

3. Mutually develop processes separate from this agreement to allow exchange of funds, should that be mutually determined to be required during the period of this CRADA

NOTE 1: The parties agree that Article 9.2 Extent of Liability, will be subject to and interpreted in accordance with Florida Statutes, Section 768.28.

NOTE 2: FOREIGN ACCESS TO TECHNOLOGY

All work described in this SOW shall be subject to the following with regard to foreign access to technology.

The Collaborators agree that research findings and technology developments under this CRADA may constitute a significant enhancement to the national defense. Accordingly, access to important technology developments under this CRADA by foreign firms or institutions must be carefully controlled. These controls are in addition to, and are not intended to change or supercede, the provisions of the International Traffic in Arms Regulation (22 CFR pt. 121 et seq.) the DoD Industrial Security Regulation (DoD 5220.22-R) and the Department of Commerce Export Regulation (15 CFR pt.770 et seq.)

The non-Navy Collaborator shall provide timely notice to the government's designated Principal Investigator of any proposed transfer of research findings and technology developments under this CRADA to foreign firms or institutions. If the government determines that the transfer may have adverse consequences to the national security interests of the United States, the non-Navy Collaborator and the government shall jointly endeavor to find alternatives to the proposed transfer which obviate or mitigate potential adverse consequences of the transfer but which provide substantially equivalent benefits to the non-Navy Collaborator.

APPENDIX B

CONFIRMATORY LICENSE AGREEMENT	*1. APPLICATION FOR (Title of Invention)*
2. INVENTOR(S) AND AFFILIATION	
3. PATENT APPLICATION SERIAL NO.	*4. PATENT APPLICATION FILING DATE*
5. NAVY ACTIVITY (Name, address, point of contact)	*6. NON-NAVY ACTIVITY (Name, address, point of contact)*
7. CRADA AGREEMENT NO.	*8. DATE OF THIS AGREEMENT*

9. The Invention identified above is a "Subject Invention" under Article 7 Intellectual Property included with the CRADA identified in Box 7 between the Department of the Navy and Non-Navy Activity identified in Box 6.

This document is confirmatory of the nonexclusive, irrevocable, paid-up license to practice the identified Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the receiving party, and of all other rights acquired by the receiving party by the referenced clause.

This license is granted to

_____ *the Government*

(Select one)

_____ *Non-Navy Activity identified in Box 6*

under this CRADA in the identified Invention, Patent Application and any resulting patent.

The licensee is hereby granted an irrevocable power to inspect and make copies of the above-identified Patent Application.

_____ *ACTIVITY NAME OF LICENSOR*

_____ *SIGNATURE*

_____ *NAME (Typed or Printed)*

_____ *TITLE*

_____ *BUSINESS TELEPHONE*

University of Central Florida

Multi-Disciplinary University Research Initiative

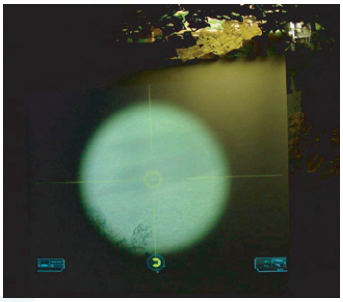
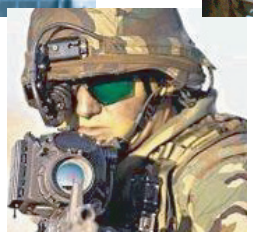
MURI-OPUS

Operator Performance Under Stress

www.mit.ucf.edu

VOLUME TWO




2002 Executive Summary

Table of Contents

NCRADA

A Collaborative Partnership between:

University of Central Florida

Contact Information:

University of Central Florida • MIT² Laboratory

3100 Technology Pkwy, Suite 337 • Partnership II • Orlando, Florida 32826-0544

http://www.mit.ucf.edu

Executive Summary

Accomplishments to Date

We have now reached the end of the first year of the Multiple University Research Initiative (*MURI*) program on Operator Performance Under Stress (*OPUS*). This report represents an account of progress to date, which contains numerous successes achieved at this very early stage. Following the original proposal, we have elaborated upon a number of areas that remain unresolved about the action of stress on performance. Each of these issues has served as a central theme in our on-going series of 'White Papers' that act to guide empirical research efforts. The two completed 'White Papers' are included in the present report and are to be followed by three additional 'White Papers' that, added to our initial model, will form the five pillars of our theoretical foundation. Although we have detailed the progress of each discrete section of the MURI program individually, we continue to strengthen and exploit each of the theoretical, methodological, and electronic links between our respective laboratories. At the University of Central Florida, the MIT^2 research group has made considerable progress. We have been able to identify, generate, and acquire research test beds to evaluate our respective theoretical propositions. From this first year's work we now know much more about variations of spatio-temporal perception under stressful conditions. The Team Performance Laboratory (*TPL*) at the University of Central Florida has investigated adaptability of teams under time pressure and high workload, and are currently using a military-based task (*Flashpoint*) to further investigate the impact of stress on team processes. MURI researchers of the Human Factors Research Laboratory (*HFRL*) at the University of Minnesota developed the *VIKING* experimental environment to assess performance on tasks that impose information processing demands similar to those faced by soldiers in combat environments. Further development work on the *VIKING* platform is underway at Catholic University, the University of Toronto, and the MIT^2 laboratory at the University of Central Florida. Minnesota researchers have also used the ANAM battery to assess the effects of physical workload on cognitive performance, finding a tendency for faster less accurate responding at high exercise levels, but much stronger effects for response time than for accuracy. The recently developed ARES battery has been used to assess the changes in cognitive performance of soldiers in the field during simulated combat, finding that many soldiers tolerate such field conditions with minimal impairment of cognitive performance if they know what is expected but that high demand events are still associated with cognitive performance changes. In addition, it appears that different forms of cognitive task are not affected equally in the context of combat. MURI researchers at Kansas State University have investigated effects of trust on team decision-making. These efforts have resulted in the identification of a viable research paradigm for the investigation of trust and team problem solving and represents one of the myriad examples of cross Laboratory technology transfer and protocol integration. Further, the KSU group has conducted two laboratory experiments on the effects of social interaction on performance. Specifically, they have investigated team problem solving and cooperative performance in distributed teams. Following our proposed form of interaction, one member of the KSU group is now working at the MIT^2 laboratory. MURI researchers in the Cognitive Science Laboratory at the Catholic University of America have completed their initial studies on the impact of automation on performance in decision-making tasks of the kind typically faced by military personnel in command and control settings. They found that as a result of complacency, automation with high (but not perfect) reliability does not always support performance. They also found that automation unreliability has a greater

performance cost for decision automation than for information automation. They are currently working on further development of the *VIKING* platform to investigate the effects of automation support on performance with tasks requiring decision-making or information acquisition and integration. In addition, they are exploring the effectiveness of adaptive automation using real-time measurement of workload under conditions of sustained workload and perceived stress. During the first year we have made extensive contacts with other Institutions and have established their cooperation with our program. Principal among these is the University of Cincinnati that has already begun value-added work on the issue of impact of workload transitions on performance. Researchers at the University of Toronto are exploring the use of wearable computer systems in arduous conditions. They have already secured funding from the Canadian Army to generate a variation of the *VIKING* task developed at the University of Minnesota. Finally, the Florida Institute of Technology has also contributed work on adaptive interface design that has had impact on the stress and design aspects represented in one of the draft papers included here. In another important step, the MURI group has signed a Cooperative Research and Development Agreement (*CRADA*) with the Naval Air Warfare Center, Training systems Division to work on mutual efforts concerning small arms development and testing in simulation environments. A focus of the MURI work is improving small arms design and capability under conditions of operational stress. Thus, this agreement facilitates the development of a high-fidelity test environment to enhance our current capabilities and permits the highest possible level of recording of response capability. In respect of all these achievements, MURI has supported over thirty researchers, including two post-doctoral researchers and over twenty graduate student researchers across all participating Institutions.

Contributions to Scientific Understanding

We have made several fundamental contributions to the scientific understanding of stress effects. We have begun to elucidate the effects of diverse sources of real-world stress on spatio-temporal perception. Using stress sources that will impact further development of the IMPRINT model, we have begun to confirm our hypotheses on the 'narrowing' of the perceptual world of the stressed operator. These efforts have burgeoned into display design guidelines for such operations as a deliverable product, a draft of which is included here. We have made substantive theoretical progress on the question of adaptive interfaces for stressful operations. In particular, we have formulated a number critical questions about the efficacy of design and the possibility of optimal design. These developments have helped characterize and strengthen currently tested empirical propositions that promise to save significant dollars against a vision of endless, iterative search for design improvement. Our progress has been facilitated by naturalistic field observations and evaluations. The opportunity to conduct real-world test and evaluation on stressed soldiers accords with our theoretical proposition that a fully articulated theory can only emerge from an understanding of actual operations. While we continue to pursue laboratory and simulation efforts, we look to a union of forms of observation against which to test hypotheses formulated in the respective 'White Papers.' Our experiments on combined physical and cognitive effort has resulted in a surprising and important discovery. We call this the 'table-top' tolerance limit phenomenon. This phenomenon indicates that even when stress levels are significantly high the performance remains relatively constant to a point of sudden incipient failure. This finding has significant implications for both our future work and planning for military operations in general. Identifying symptomatic characteristics at the edge of failure (e.g., progressive increase in performance variability) may provide a predictive marker that enables the identification of failing operators. This would prove to be a crucial theoretical and practical

discovery. We are also now beginning to distill ways in which team performance under stress is not simply the product of individual contributions. In particular, we are starting to show how workload sharing and task division and reallocation by team units acts to diffuse unacceptably high levels of demand. We are using this insight as a basis for designs in adaptive interfaces for stress in which demand may be shared between exposed soldiers and support personnel in remote locations. Initial developments in our collective work demonstrate how direct monitoring of operator state may be mapped into task resolution, well beyond the traditional conception of 'adaptive' interfaces. These new, neuro-ergonomic insights present the next generation level of conceptions for human-machine interaction. We have also found that stress influences mutual trust levels in combined decision-making conditions. The continuation of this effort seeks to explicate the specific mechanism of stress-induced deterioration. These represent only a selection of insights derived from the contained work, others of which are articulated in the specific papers.

Research for the Coming Year

In the coming year we look to complete the three 'White Papers' as our theoretical efforts look toward reaching fruition. Programmatic research evaluations related to the propositions set forth in the white papers will be pursued. Further experimentation is planned on the effects of time-pressure, cognitive workload change, distracting pain, degree of operator's control over the task environment, and noise stress on a variety of performance capacities. This experimentation permits tests of our theoretical models of the spatial and temporal components of stress and task performance, as well as the recruitment, allocation, and depletion of resources under conditions of high workload and stress. Further, we plan further evaluation of target acquisition capacities and decision strategies as tests of fuzzy signal detection under stress. Finally, the characteristics of operators that predict variation in performance will be evaluated in each of these studies. Such knowledge will inform development of selection and training procedures, as well as identify characteristics of individuals that render them more stress resistant. We are in the process of constructing a descriptive matrix that will inform current mechanisms for predicting task performance under stress (i.e., *IMPRINT*). In light of the recent integration of IMPRINT with ACT-R, future efforts will evaluate stress-related performance deviations from modeled data under standardized conditions. Specifically, the ensuing database and meta-analytic review will aid in modifying and creating new degradation algorithms for current and future stressors within IMPRINT and mold the existing task taxonomy that is central to the effectiveness of performance shaping functions used within the task-network model. Each of the proposed empirical studies is derived from a confluence of three constraints. First, through reference to our growing matrix, we ensure that such work is not already present in the scientific literature. Second, we explicitly distill how such experimental work addresses one or more of the facets of theoretical uncertainty as presented in the foundational 'White Papers.' Finally, we will examine how our research on stress can be integrated into the IMPRINT/ACT-R modeling architectures. Our articulation of a complete theory and its instantiation in various modeling efforts is clearly contingent upon a much better understanding of the spectrum of influences experienced by actual soldiers. We shall continue our field efforts in order to understand and incorporate the nuances that the real-world presents. Further, our field researchers help us identify those Forces personnel who would be the recipients and beneficiaries of our collective efforts. In laboratory testing, we have already started to evaluate the effects of even higher levels of combined physiological and cognitive stressors while our synthetic and empirical evaluations of subtle performance variations from the 'edge of intolerance' is proceeding on a number of fronts.

Table of Contents

MURI Divisions

University of Central Florida (Peter Hancock)

Supported Work

Using a Virtual Combat Simulator to Examine Soldier Performance in Infantry Combat *Poster* 110th Annual Meeting of the American Psychological Association *Greenwood-Ericksen, A. Weaver, J. L. and Hancock. P. A. (2002)*

Driver Characteristics and Technology: Enhancing Transportation Safety and Efficiency *Poster* 110th Annual Meeting of the American Psychological Association *Weaver, J. L., Bradley, K.A., Thropp, J. E., Hancock, P. A., & Mouloua, M. (2002)*

Effects of Stress on Human Performance *Taylor Award Presentation* 110th Annual Meeting of the American Psychological Association *Hancock, P.A. (2002)*

Operator Performance Under Stress *Presentation* 110th Annual Meeting of the American Psychological Association *Hancock, P. A. (2002)*

Effects of context-relevant stress in PC-based game training *Poster* 110th Annual Meeting of the American Psychological Association *Morris, C.S., Shirkey, E.C., Hancock, P.A. & Mouloua, M. (2002)*

The Effect of Noise Stress during Search for Two Identical Targets in a Realistic Visual Scene *Poster* 110th Annual Meeting of the American Psychological Association *Diaz, D.D., Hancock, P.A., Sims, V.K. & Brill, C. (2002)*

A Multi-sensory Device and Protocol for Online Workload Assessment *Poster* 110th Annual Meeting of the American Psychological *Association Brill, J.C., Mouloua, M., Hancock, P.A & Gilson, R.D. (2002)*

A Multi-sensory Device and Protocol for Online Workload Assessment *Proceedings* 110th Annual Meeting of the American Psychological Association *Brill, J.C., Mouloua, M., Hancock, P.A & Gilson, R.D. (2002)*

The Influence of Stress and Individual Differences on the Perception of Duration *Paper Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association *Weaver, J.L. & Hancock, P.A. (2002)*

The Influence of Stress and Individual Differences on the Perception of Time *Paper Presentation* Mid-Year Symposium of Division 21: Applied Experimental

and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association *Weaver, J.L., Sims, V.K., Ganey, H.C.N., Hancock, P.A., Bradley, K.A. & Mouloua, M. (2002)*

Visual search for two identical targets: The role of ecological validity. *Paper Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association *Diaz, D.D., Sims, V.K., & Hancock, P.A. March, (2002)*

The effect of time stress on mission performance in a virtual combat mission. *Paper Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association *Greenwood-Ericksen, A., Ganey, H.C.N., Weaver, J.L., Mouloua, M., & Hancock, P.A. (2002)*

Training and performance outcome: Advanced technologies, realistic graphics, and increased simulation may not be contributors. *Paper Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association *Morris, C.S., Tarr, R.W., Shirkey, E., Hancock, P.A., & Mouloua, M. (2002)*

Is there a Science of (Adaptive) Interface Design for Decision-Making? *Proceedings* 23rd Army Science Conference *Hancock, P.A. (2002)*

Performance Under Stress: A Re-evaluation of a Foundational Law of Psychology *Proceedings* 23rd Army Science Conference *Hancock, P.A., Ganey, H.C.N. & Szalma, J.L. (2002)*

Off-the-self Games and Training: Lessons Learned *Proceedings* 23rd Army Science Conference *Morris, C.S. & Hancock, P.A. (2002)*

Stress and Human Information Processing: A Descriptive Framework Presented in a Novel Manner *Proceedings* 23rd Army Science Conference *Hancock, P.A., Ward. P., Szalma, J.L., Stafford & Ganey, H.C.N. (2002)*

Stress and Human Information Processing: A Descriptive Framework Presented in a Novel Manner *Poster* 23rd Army Science Conference *Hancock, P.A., Ward. P., Szalma, J.L., Stafford & Ganey, H.C.N. (2002)*

The Point of Diminishing Immersive Return: Implications for Simulation-based Training *Proceedings* 23rd Army Science Conference *Morris, C.S., Ganey, H.C.N., Ross J. & Hancock, P.A. (2002)*

Sans Subjectivity, Ergonomics is Engineering *Refereed Publication*

Hancock, P.A., Weaver, J.L. & Parasuraman, R. (2002). Sans Subjectivity, Ergonomics is Engineering, *Ergonomics*, **45** (14), 991-994.

Workload and Stress of Configural Displays in Vigilance Tasks *Proceedings* 46th Annual Human Factors and Ergonomics Society *Szalma, J.L. (2002)*

Individual Differences in the Stress and Workload of Sustained Attention *Proceedings* 46th Annual Human Factors and Ergonomics Society *Szalma, J.L. (2002)*

Individual Differences in the Stress and Workload of Sustained Attention *Poster* 46th Annual Human Factors and Ergonomics Society *Szalma, J.L. (2002)*

The Impact of Stress on Adaptability in Teams *Poster* 46th Annual Human Factors and Ergonomics Society *Priest, H. A., Burke, S., Munim, D. & Salas, E. (2002)*

Applications of Fuzzy Signal Detection Theory to Responses to Terrorism *Poster* 46th Annual Human Factors and Ergonomics Society *Szalma, J.L., Hancock, P.A., Mouloua, M., & Parasuraman, R. (2002)*

A Human Factors Approach to the Investigation of Stress *Proceedings* Annual Southeastern Psychological Association *Ganey, H.C.N., Hancock, P.A., Weaver, J.L., Greenwood-Erickson, A. & Mouloua, M. (2002)*

The Distortion of Perceptual Space-time Under Stress *White Paper Hancock, P.A., Szalma, J.L. & Weaver, J.L. (2002)*

2002 Team Performance Laboratory's Update Report
Bowers, C., Salas, E., Asberg, K., Burke, S., Priest, H. & Milham, L. (2002)

Combat readiness and stress: Laboratory investigations of teams *Presentation* 110th Annual Meeting of the American Psychological Association *Bowers, C., Salas, E., Asberg, K., Burke, S., Priest, H. & Milham, L. (2002)*

On Mental Resources and Performance under Stress *White Paper Szalma, J.L. & Hancock, P.A. (2002)*

Facilitated Work
The Triples Rule *Refereed Publication*
Hoffman, R.R., Hayes, P., Ford, K.M. & Hancock, P.A. (2002). *IEEE Intelligent Systems*, **17** (3), 62-65

Catholic University (Raja Parasuraman)

Supported Work

Effects of Information and Decision Automation Multitask Performance *Paper* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association *Rovira, E., Zinni, M. & Parasuraman, R. (2002)*

Effects of Unreliable Automation on Decision Making in Command and Control *Paper* 46th Annual Meeting of the Human Factors and Ergonomics Society *Rovira, E., McGarry, K. & Parasuraman, R. (2002)*

Adaptive Automation Matched to Human Mental Workload *Presentation* NATO Advanced Research Workshop on Operator Functional Assessment and Impaired Performance *Parasuraman, R. (2002)*

Quarterly Update 03/2002

Quarterly Update 07/2002

University of Minnesota (Wayne Harris)

Supported Work

The Effect of Physical Workload on Cognitive Performance *Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association *Harris, W.C., Walker, A.J., Leon, A.S. & Hancock, P.A. (2002)*

Infantry Cognitive Changes during Simulated Combat *Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association *Harris, W.C. (2002)*

VIKING: Development of an infantry scenario multi-task environment. *Presentation* Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association *Harris, W. C., Parasuraman, R., Zinni, M., & Hancock, P. A. (2002)*

VIKING training program: A technical report. *Presentation* A technical report on VIKING from UM *Harris W C (2002)*

Stress and cognitive performance field assessment – Quarterly Report 12/2002 *Harris, W. C.*

Stress and cognitive performance field assessment – Quarterly Report 03/2002 *Harris, W. C.*

Stress and cognitive performance field assessment – Quarterly Report 09/2002 *Harris, W. C.*

Phase I – Progress report 03/2002 *Leon, A.J.*

Kansas State (Kip Smith)

Supported Work

The Effect of Trust on Team Problem Solving *Presentation* Oklahoma-Kansas Society for Judgment and Decision Making *Fosse, N., Murphy, L., Perry, J., Ratcliff, L. & Smith, K. (2002)*

The Impact of an Economic Manipulation of Trust on Team Problem Solving *Poster* 46th Annual Human Factors and Ergonomics Society *Swan, B. Shane, A., Ratcliff, E., Pruitt, N., Kearney, A., Freund, S., Fosse, N., & Smith, K. (2002)*

Progress report - Smith, K. 03/2002

Progress report - Smith, K. 06/2002

Progress report - Smith, K. 09/2002

Progress report - Smith, K. 12/2002

Florida Institute for Technology (John Deaton)

Facilitated Work

Progress Report - Homes Regional Medical Center First Flight Air Ambulance Evaluation Deaton, J. (2002)

University of Cincinnati (Joel Warm)

Supported Work

Progress report - Warm, J. 09/2002



Using a Virtual Combat Simulator to Examine Soldier Performance in Infantry Combat

A. Greenwood-Ericksen, J.L. Weaver, and P. A. Hancock.
University of Central Florida

ABSTRACT

Traditionally, it has been extremely difficult to examine soldier performance in infantry combat environments because of the extreme expense and risk to personnel involved in large scale practice operations. The present study addresses the issue of stress and soldier performance, but does so using a virtual combat simulation based on commercial off the shelf (COTS) software. In infantry combat, soldiers routinely operate in highly stressful conditions while simultaneously making decisions under time pressure.

The aim of this ongoing study is to address the effect of time stress in a virtual soldier simulator with the eventual goal of comparing it to existing stress data on soldiers in traditional training exercises. If evidence can be found that responses to time stress are similar in virtual and real-life training environments, then it might be possible to conduct future soldier performance studies in virtual environments while maintaining a high level generalizability to combat performance.

Forty undergraduate subjects were trained in the use of a commercial off the shelf video game (COTS) soldier simulator and then completed a simulated prisoner of war (POW) rescue mission under one of two time-pressure conditions. In the control condition participants experienced no time stress while completing the scenario, while participants in the experimental condition performed the same mission under time pressure. Observed effects of mission success or failure and previous video game ability on the manipulation of time pressure and performance in a simulated virtual combat environment are discussed.

INTRODUCTION

Combat is one of the most stressful tasks performed by human beings, from both a physical and a physiological perspective (Grossman, 1996). Moreover, the consequences of inadequate performance in this role can be catastrophic not only for the individual soldier, but for the unit, army, or even the nation as a whole. Because of its inherently stressful nature and the vital importance of successful outcomes in combat, it is vital that an experimental examination of factors affecting basic soldier performance be performed in order to gain a more complete picture of soldier performance under stress. To this end, a between subjects study was conducted to investigate participant performance on a low-fidelity soldier simulator under varying levels of time stress. Time stress has been previously established as a valid source of stress (Hancock & Desmond, 2001, Doob, 1971)

The purpose of the current study was to examine the effects of perceived time pressure on the performance of soldier operators engaged in a simulated mission environment. We predicted that certain indices of individual performance would be degraded under conditions of significant time-pressure (the IV). Dependent variables included a series of subjective self-report assessments of the task itself and their engagement in it, as well as the speed and success of mission completion. Specifically, we expected participants in the experimental condition to experience more subjectively-rated time pressure, take longer to complete the task, and rate the task as more difficult and their own performance as poorer. In addition, we predicted that those who successfully rescued all the POWs would rate their own performance more highly, as would those who rated themselves as better in global video game ability.

The larger purpose of this study was to produce data on soldier performance under time stress which might eventually be integrated into a larger body of research to develop a dynamic model of stress and soldier performance in combat. Through a more complete understanding of the factors influencing performance in a virtual combat simulator it might be possible both to improve training methods and increase soldier performance in actual combat.

METHODOLOGY

Participants:

Participants were 41 male UCF undergraduates

Equipment:

The primary platform for this task is a commercially-available video game called Delta Force: Land Warrior, developed by Novalogic inc. This platform was chosen for its face validity and its wide acceptance across various groups involved in the Army MURI-OPUS (Multi-disciplinary University Research Initiative-Operator Performance Under Stress) project.

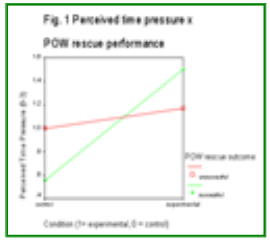
Experimental Procedure:

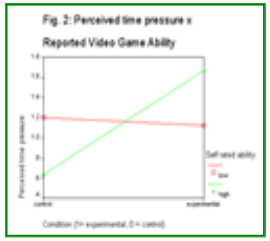
Control: In the control condition, participants performed mission objectives without consequences for lack of speed in accomplishing said objectives.

Experimental: In the experimental condition, participants were told that accomplishing the mission objectives would be impossible upon expiration of the time, and that speed was essential. It was expected that participants assigned to this condition would experience significant time-pressure in carrying out the task.

Training: Participants were trained to use the simulation by completing three missions, averaging about 15 minutes in length during which they were trained by the experimenter on the various skills and strategies necessary to complete the mission.

They were then exposed to a mission identical to one of the training scenarios. Their objective in this scenario was to successfully rescue as many hostages as possible from a POW camp. Successful mission accomplishment was defined as rescuing all eight POWs. Participants completed the rescue mission without online feedback from the experimenter and their performance was digitally recorded to allow for later coding and analysis by raters. Coding reliability was 100% across ratings.





RESULTS

Manipulation check:

A main effect of condition on perceived time pressure was observed such that participants in the time pressure condition (M =1.333, SD =.571) rated themselves as experiencing more time pressure than those in the control group (M =.778, SD =.616), $F(1,36) = 9.110$, $p = .005$, $\eta 2 = .202$.

Main effects:

A main effect of self-rated video game ability on perceived task performance was observed such that participants who rated themselves as low in video game ability believed that they didn't perform as well (M = 1.963, SD = .157) on the task as those who reported themselves as being high in video game ability (M = 2.477, SD = .140), $F(1,36) = 6.858$, $p < .05$, $\eta 2 = .160$.

There was also a main effect of mission success (rescuing all 8 POW's) on perceived task performance such that participants who successfully completed the mission objective of rescuing 8 POW's (M = 2.588, SD = .507) felt that they had performed better than those who did not rescue all the POWs (M = 2.130, SD = .548), $F(1, 36) = 6.856$, $p < .05$, $\eta 2 = .160$.

Interactions:

A significant interaction was observed between time pressure condition and mission performance level (all POW's rescued or not) on perceived time pressure, $F(1,36) = 4.464$, $p < .05$, $\eta 2 = .110$, such that successful mission completers in the experimental condition (M = 1.500, SD = .203) experienced significantly higher time pressure than those in the control group (M =.556, SD = .192), while unsuccessful mission completers in the experimental group (M = 1.167, SD = .166) did not experience significantly different time pressure than those in the control group (M = .727, SD = .155) according to Tukey's post-hoc comparisons (Fig 1).

Another significant interaction was observed between time pressure condition and self rated video game ability on perceived time pressure, $F(1,36) = 5.707$, $p < .05$, $\eta 2 = .180$, such that high ability participants in the experimental group (M = 1.667, SD = .244) experienced significantly higher time pressure than high ability participants in the control group (M = .636, SD = .183), but low ability subjects in the experimental group (M = 1.125, SD = .214) did not significantly differ from low ability subjects in the control group (M = 1.200, SD = .271) according to Tukey's post-hoc comparisons (Fig 2).

No additional significant main effects or interactions were observed on perceived time pressure, task difficulty, task performance, or on elapsed mission time.

DISCUSSION

As seen above, the manipulation check indicates successful manipulation of perceived time pressure across all participants. This finding confirms that participants did in fact perceive more time pressure in the experimental than the control condition. This was an initial concern as the pilot data on this study failed to produce a significant difference as a function of our manipulation. Although we did find a self-reported effect of time pressure manipulation, we were unable to identify a main effect of time pressure on any of our behavioral measures. However, the presence of significant interactions with perceived time pressure as a dependent variable appears to indicate a more complex relationship between performance measures and time stress.

In addition, several of the observed performance effects, although significant, appear to be small as indicated by eta squared values in the .16- .20 range. One potential explanation for our failure to find performance effects as a function of our time pressure manipulation might relate to the relatively restricted range of time pressure perceived by groups in both conditions. In other words, although the groups did differ significantly, participants in the experimental group rated themselves as experiencing a time pressure level only around 1.3 on a scale of 0-3 (M=1.333, SD =.571), and only .78 in the control group (M=.778, SD =.616), indicating that not much time pressure was perceived in either group in relation to the overall scale of measurement. These results suggest that a stronger manipulation of time pressure might have improved the size of these effects.

In general, the data support several conclusions. First, as discussed above, our manipulation check (perceived time pressure) indicates that participants in the experimental condition did feel more pressured than those in the control group.

The presence of a significant main effect of success versus failure on the primary mission goal of retrieving all POW's alive on perceived task most likely indicates that participants who lost more POW's felt that they performed more poorly.

DISCUSSION (cont.)

In addition, the observed main effect of video game ability on perceived task performance indicates that participants with low self-reported video game ability believed that they performed significantly more poorly than those with high video game ability. Since this questionnaire item probed self perception of general video game ability rather than actual objectively-rated performance, it can be argued that it could be more properly seen as a measure of global video game self-efficacy. Regardless, this finding clearly indicates an individual difference variable related to video game ability which could be responsible for differential perception of task performance.

However, it should be noted that the mean self-rating of even low ability participants was almost 2 (1.96) on a three to one scale, indicating that responses on this measure were generally high.

A significant interaction was observed between the number of POW's rescued and condition on perceived time pressure experienced by participants. Participants who successfully rescued all POW's in the experimental group experienced a significantly higher level of time pressure than those in the control group who successfully rescued all POWs. No significant difference, however, was observed between the experimental and control groups among participants who did not successfully rescue all POW's.

This finding seems to indicate a heretofore unidentified variable at work that distinguishes high and low performers on POW rescue on the basis of perceived time pressure, regardless of time pressure manipulation condition. Accordingly, it seems probable that this variable might be an individual difference variable. Past research on individual difference variables (Weaver, Bowers, and Morgan, 1996a) has shown that individuals high in self-control respond better to time based cues than those low in self-control. Therefore, self control or similar individual difference factors might be responsible for the observed effects.

A second observed interaction between condition and self-rated video game ability on perceived time pressure indicated that participants who rated themselves as high in video game ability differed significantly between the control and experimental groups but no such effect was observed among participants who rated themselves as being poor at video games. This finding also seems to indicate an individual difference variable at work in the study. Several previous studies have addressed the role of self-efficacy in moderating stress in task performance (Bandura, 1977, Bowers, Weaver, & Morgan, 1996b). Since, as argued above, our video game question addressed self perception of ability rather than actual ability it seems reasonable to interpret it as a viable measure of task-specific self-efficacy. Therefore, it seems possible that an important factor in simulated infantry task performance is not only competence with video simulations in general, but *perceived* competence in video simulations. Additionally, the role of time-pressure induced affective state is likely to alter still further the responses and reactions of participants engaged in these types of tasks.

Other dependent variables failed to show significant effects. No significant effects or interactions were observed on mission completion time or reported on the subjective measure of task difficulty. Additionally, main effects of time pressure condition on mission completion time, perceived task difficulty, and perceived task performance were predicted but not observed, indicating that, contrary to our hypotheses, time pressure manipulation did not seem to have any effect on the speed with which participants completed the task, nor on how hard they found it or how well they thought they performed.

REFERENCES

Allan, L.G. (1979) The perception of time. *Perception and Psychophysics, 26,* 340-354.

Allender, L., Lockett, J., Kelley, T.D., Salvi, L., Mitchell, D., Headley, D.B., Promisel, D., Richer, C., & Feng, T. (1995). Verification, validation, and accreditation of a soldier-system modeling tool. *Proceedings of the Human Factors and Ergonomics Society, 39,* 1219-1223.

Doob, L.W. (1971) *Patterning of Time.* .New Haven: Yale University Press.

Fatkin, L.T., King, J.M., & Hudgens, G.A. (1990). Evaluation of stress experienced by Yellowstone Army firefighters (Tech. Mem. No. 9-90). U.S. Army Research Laboratory, Human Research and Engineering Directorate. Aberdeen Proving Ground, MD.

Grossman, D. (1996). *On killing: The psychological cost of learning to kill in war and society.* New York: Little, Brown and Company.

Hancock, P.A., & Desmond, P.A. (Eds.). (2001). *Stress, workload and fatigue.* Lawrence Erlbaum, Mahwah: NJ.

Maule, A.J., Hockey, G.R., Bdzola, L. (2000). Effects of time pressure on decision-making under uncertainty: Changes in affective state and information processing strategy. *Acta Psychologica, 104,* 283-301.

Stanney, K., Salvendi, G. Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *International Journal of Human-Computer Interaction, 10,* 135-187.

ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Driver Characteristics and Technology: Enhancing Transportation Safety and Efficiency

J. L. Weaver, K.A. Bradley, J. E. Thropp, P. A. Hancock, & M. Mouloua
University of Central Florida

Introduction

As technologically advanced vehicles become increasingly prevalent in modern day transportation, it has become essential to study the interaction between telematic devices and the driver as well as how individual driver characteristics and driver error may relate to these interactions. The purpose of the current research was to serve as preliminary research regarding the relationship between driver characteristics and driver error. The ultimate goal is to then develop a "driver profile" consistent with high levels of driver error in order to identify "at risk" drivers. It is hoped that the profile would facilitate investigations regarding driver distraction and cognitions and how they might influence driver interaction with and use of telematic devices.

Efforts to address related issues have been initiated by researchers in various countries attempting to better their understanding of factors related to aggressive driving, driver error, and road rage. As a result of these efforts, a number of useful measurement instruments have been developed. For example, the Driver Behavior Inventory (DBI) assesses factors related to stress and driving (Glendon, Dorn, Matthews, Gulian, Davies, & Debney, 1993). These factors include (1) driving aggression, (2) irritation when passed, (3) driving alertness, (4) driving dislike, (5) frustration when failing to overtake, and (6) general driver stress. These factors have been found to be related to such driving behaviors as caution in overtaking and errors in following other vehicles.

The Driver Behavior Questionnaire (DBQ) assesses driving errors and violations (Parker, Reason, Manstead, & Stradling, 1995). This instrument is valuable for classifying the types of errors made and the circumstances under which these errors are most likely to be exhibited. Another approach to studying this phenomenon has been limited research regarding differences in attributional styles or driver attitudes. Specifically, research has explored the attributions made by actors (drivers) versus observers for particular vehicular incidents (Herzog, 1994). Specifically, drivers attribute blame for accidents differently if they are at fault vs. situations when another driver is at fault. The Cognitive Failures Questionnaire (CFQ) measures cognitive errors such as slips and lapses (Broadbent, Cooper, FitzGerald, & Parkes, 1982). Cognitive failures are defined as perceptual, memory, or action failures in which events are not congruent with intention (Broadbent et al., 1982). The current research explores how attributions are related to various types of cognitive errors and the degree that they relate to identifiable driving errors. Thus, the following hypotheses were developed.

Hypotheses

It was hypothesized that attributional complexity (as measured by the Attributional Complexity Scale) would correlate negatively with: the violations, lapses, and errors clusters of the DBQ; the aggression, dislike of driving and related anxiety, and irritation when passed clusters of the DBI; with number of accidents, and the Driver Anger Scale.

It was also hypothesized that cognitive failures (as measured by the Cognitive Failures Questionnaire) would correlate positively with: the errors, lapses, and violations clusters of the DBQ; the general driver stress, aggression, dislike of driving and related anxiety, irritation when passed clusters of the DBI; and number of accidents. A negative correlation between cognitive failures and the driving alertness cluster of the DBI was also hypothesized.

It was also hypothesized that: the violations cluster of the DBQ would correlate positively with number of accidents; that the dislike of driving and related anxiety and driver aggression clusters of the DBI would correlate positively with the lapses, errors, and violations clusters of the DBQ; and that the driver alertness cluster of the DBI would correlate negatively with the lapses, errors, and violations clusters of the DBQ all in replication of prior findings.

Method

The participants for this study consist of 417 undergraduate psychology students attending the University of Central Florida. Participants were compensated for their time with extra course credit. The researchers recruited undergraduate psychology students from their psychology courses. Participants were given a consent form to sign followed by a packet of questionnaires which include the DBI, the DBQ, the Attributional Complexity Scale, the Driver Anger Scale (Deffenbacher, Oetting, & Lynch, 1994), and the CFQ (sample questionnaire items can be found in Table 1). Participants then received a debriefing sheet including researcher contact information. No foreseeable risks were expected as a result of participating in this study.

Table 1. Sample Questionnaire Items

Driving Anger Scale

Someone is driving too slowly in the passing lane holding up traffic.
Someone pulls right in front of you when there is no one behind you.

Attributional Complexity Scale

I don't usually bother to analyze and explain people's behavior.
I am very interested in understanding how my own thinking works when I make judgements about people or attach causes to their behavior.

Cognitive Failures Questionnaire

Do you read something and find you haven't been thinking about it and must read it again?
Do you find you forget people's names?

Driving Behavior Inventory

It annoys me to drive behind slow moving vehicles.
I feel bothered when passed at an intersection.

Driver Behavior Questionnaire

Drive especially close to the car in front as a signal to its driver to go faster or get out of the way.
Realize that you have no clear recollection of the road along which you have just been traveling.

Table 2. Correlation Matrix

Correlations

		DBQ cluster 1: lapse	DBQ cluster 2: error	DBQ cluster 3: violation	Cognitive failures questionnaire total	attributional complexity scale total	aggressive driving scale total	DBI cluster 1: driving aggression	DBI cluster 2: dislike of driving and related anxiety	DBI cluster 3: driving alertness	DBI cluster 4: irritation when overtaken (passed)	DBI cluster 5: overtaking (passing) tension	DBI cluster 6	Number of moving violation tickets received for accidents	Number of accidents involved in in the last three years	Estimated mileage for the last three years (compared to an average of 12,000 miles per year)
DBQ cluster 1: lapse	Pearson Correlation	1.000	.805**	.598**	.556**	-.213**	.153	.381**	.313**	.313**	.266**	.139	.352**	.196*	.007	.054
	Sig. (2-tailed)		.000	.000	.000	.000	.063	.000	.000	.000	.001	.083	.000	.044	.235	.506
	N	160	160	156	157	149	148	160	158	159	158	156	155	106	153	157
DBQ cluster 2: error	Pearson Correlation	.805**	1.000	.681**	.459**	-.204**	.217**	.444**	.297**	.243**	.340**	-.029	.435**	.148*	.041	.024
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.566	.000	.016	.003	.828
	N	160	408	401	402	394	389	408	404	407	406	402	401	262	395	404
DBQ cluster 3: violation	Pearson Correlation	.598**	.681**	1.000	.382**	-.163**	.294**	.709**	.264**	.055	.306**	-.168**	.635**	.123*	.144**	.103*
	Sig. (2-tailed)	.000	.000		.000	.001	.000	.000	.000	.273	.000	.001	.000	.049	.004	.040
	N	156	401	402	397	388	383	402	398	401	400	397	396	259	391	399
Cognitive failures questionnaire total	Pearson Correlation	.556**	.459**	.382**	1.000	.009	.276**	.358**	.327**	.036	.164**	-.001	.405**	.015	.104*	.040
	Sig. (2-tailed)	.000	.000	.000		.866	.000	.000	.000	.908	.001	.991	.000	.806	.039	.423
	N	157	402	397	409	395	390	409	405	408	407	404	403	262	394	403
attributional complexity scale total	Pearson Correlation	-.213**	-.204**	-.163**	.009	1.000	-.044	.098	.010	.256**	.050	-.029	.038	.009	.065	.197**
	Sig. (2-tailed)	.000	.000	.001	.866		.388	.054	.839	.000	.317	.564	.447	.885	.203	.003
	N	149	394	388	395	401	384	401	398	400	399	395	394	256	385	394
aggressive driving scale total	Pearson Correlation	.153	.217**	.294**	.276**	-.044	1.000	.447**	.403**	.235**	.437**	-.075	.604**	.015	-.014	.013
	Sig. (2-tailed)	.063	.000	.000	.000	.388		.000	.000	.000	.000	.138	.000	.807	.782	.794
	N	148	389	383	390	384	395	395	392	394	393	389	388	253	380	389
DBI cluster 1: driving aggression	Pearson Correlation	.381**	.444**	.709**	.358**	-.098	.447**	1.000	.306**	.223**	.360**	.242**	.604**	.033	.101*	.141**
	Sig. (2-tailed)	.000	.000	.000	.000	.054	.000		.000	.000	.000	.000	.000	.597	.044	.004
	N	160	408	402	409	401	395	416	412	415	414	410	409	265	400	408
DBI cluster 2: dislike of driving and related anxiety	Pearson Correlation	.313**	.297**	.264**	.327**	.010	.403**	.306**	1.000	.160**	.368**	.126*	.621**	.080	.080	.036
	Sig. (2-tailed)	.000	.000	.000	.000	.839	.000	.000		.000	.000	.010	.000	.197	.111	.474
	N	158	404	398	405	398	392	412	412	412	410	406	406	262	396	404
DBI cluster 3: driving alertness	Pearson Correlation	-.313**	-.243**	-.055	-.006	.256**	.235**	.223**	.160**	1.000	.072	-.225**	.224**	-.127*	.012	.139**
	Sig. (2-tailed)	.000	.000	.273	.908	.000	.000	.000	.000		.146	.000	.000	.039	.816	.005
	N	159	407	401	408	400	394	415	412	415	413	409	409	264	399	407
DBI cluster 4: irritation when overtaken (passed)	Pearson Correlation	.266**	.340**	.306**	.164**	-.050	.437**	.360**	.368**	.072	1.000	.065	.600**	.131*	.149**	.093
	Sig. (2-tailed)	.001	.000	.000	.001	.317	.000	.000	.000	.146		.189	.000	.034	.003	.945
	N	158	406	400	407	399	393	414	410	413	414	410	409	264	398	406
DBI cluster 5: overtaking (passing) tension	Pearson Correlation	.139	-.029	.168**	-.001	-.029	-.075	-.242**	.126*	-.225**	.065	1.000	.034	.047	.041	-.013
	Sig. (2-tailed)	.083	.566	.001	.991	.564	.138	.000	.010	.000	.189		.490	.445	.418	.797
	N	156	402	397	404	395	389	410	406	409	410	410	409	261	394	402
DBI cluster 6	Pearson Correlation	.352**	.435**	.635**	.405**	-.038	.604**	.604**	.621**	.224**	.600**	.034	1.000	.110	.133**	.104*
	Sig. (2-tailed)	.000	.000	.000	.000	.447	.000	.000	.000	.000	.000	.490		.076	.008	.038
	N	155	401	396	403	394	388	409	406	409	409	409	409	260	393	401
Number of moving violation tickets received for accidents	Pearson Correlation	.196*	.148*	.123*	-.015	.009	.015	.033	.080	-.127*	.131*	.047	.110	1.000	.305**	-.026
	Sig. (2-tailed)	.044	.016	.049	.806	.885	.807	.597	.197	.039	.034	.445	.076		.000	.672
	N	106	262	259	262	256	253	265	262	264	264	261	260	265	259	265
Number of accidents involved in in the last three years	Pearson Correlation	.007	.149**	.144**	.104*	.065	.014	.101*	.080	.012	.140**	.041	.133**	.305**	1.000	.034
	Sig. (2-tailed)	.235	.003	.004	.039	.203	.782	.044	.111	.816	.003	.418	.008	.000		.495
	N	153	395	391	394	385	380	400	396	399	398	394	393	259	400	397
Estimated mileage for the last three years (compared to an average	Pearson Correlation	.054	.024	.103*	.040	.197**	.013	.141**	.036	.138**	.093	.013	.104*	.026	.034	1.000
	Sig. (2-tailed)	.506	.628	.040	.423	.003	.794	.004	.474	.005	.945	.797	.038	.672	.495	
	N	157	404	399	403	394	389	408	404	407	406	402	401	265	397	408

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Results

Reliability analyses were conducted for each measurement scale and for each cluster of the DBI and the DBQ. The results for each are as follows: CFQ (Alpha = .91), Attributional Complexity (Alpha = .86), Driver Anger Scale (Alpha = .93), DBI (Alpha = .87), DBI driving aggression cluster (Alpha = .83), DBI dislike of driving and related anxiety (Alpha = .68), DBI driving alertness cluster (Alpha = .71), DBI irritation when passed cluster (Alpha = .87), DBI general driver stress cluster (Alpha = .82), DBQ (Alpha = .94), DBQ lapses cluster (Alpha = .85), DBQ errors cluster
(Alpha = .86), and DBQ violations cluster (Alpha = .79).

In order to test the proposed hypotheses, Pearson product-moment correlation coefficients were calculated. As hypothesized, there were significant negative correlations between attributional complexity and the DBQ violations cluster, r (N = 388) = -.163, p < .001 (one-tailed), attributional complexity and the DBQ lapses cluster, r (N = 149) = -.313, p < .0005 (one-tailed), and attributional complexity and the DBQ errors cluster, r (N = 394) = -.294, p < .0005 (one-tailed). Contrary to our hypotheses, attributional complexity did not significantly correlate with the DBI aggression cluster, the Driver Anger Scale, the DBI dislike of driving and related anxiety cluster, the DBI irritation when passed cluster, or number of accidents.

However, also as hypothesized, there were significant positive correlations between cognitive failures and the error cluster of the DBQ, r (N = 402) = .459, p = .0005 (one-tailed), cognitive failures and the lapses cluster of the DBQ, r (N = 157) = .556, p < .0005 (one-tailed), cognitive failures and the general driver stress cluster of the DBI, r (N = 403) = .405, p < .0005 (one-tailed), cognitive failures and number of accidents, r (N = 394) = .104, p = .0195 (one-tailed), cognitive failures and the aggression cluster of the DBI, r (N = 409) = .358, p < .0005 (one-tailed), cognitive failures and the dislike of driving and related anxiety cluster of the DBI, r (N = 406) = .327, p < .0005 (one-tailed), and cognitive failures and irritation when passed cluster of the DBI, r (N = 407) = .164, p < .001 (one-tailed). Contrary to our hypotheses, cognitive failures and the driving alertness cluster of the DBI did not significantly correlate.

In support of the hypothesized replications, the violations cluster of the DBQ correlated positively with number of accidents, r (N = 391) = .144, p = .002 (one-tailed). The correlations between the dislike of driving and related anxiety cluster of the DBI and the lapses, errors, and violations clusters of the DBQ were significant and positive, r (N = 158) = .313, p < .0005 (one-tailed), r (N = 404) = .297, p < .0005 (one-tailed), and r (N = 398) = .264, p < .0005 (one-tailed), respectively. Significant positive correlations were found between the driver aggression cluster of the DBI and the lapses, errors, and violations clusters of the DBQ, r (N = 160) = .381, p < .0005 (one-tailed), r (N = 408) = .444, p < .0005 (one-tailed), and r (N = 402) = .709, p < .0005 (one-tailed), respectively. Significant negative correlations were found between the driving alertness cluster of the DBI and the lapses and errors clusters of the DBQ, r (N = 159) = -.313, p < .0005 (one-tailed), and r (N = 407) = -.243, p < .0005 (one-tailed), respectively. However, the significant correlation between the driving alertness cluster of the DBI and the violations cluster of the DBQ was not replicated.

Several other significant correlations were found that were not hypothesized. Attributional complexity correlated positively with the driving alertness cluster of the DBI (r = .256) and mileage (r = .150). Cognitive failures correlated positively with the Driver Anger Scale (r = .276). These correlations were all significant at the p < .05 level (two-tailed).

All correlations can be found in Table 2.

Discussion

The results of this study indicate that persons who are more attributionally complex tend to commit fewer driving violations than those who are less so. The results also indicate that persons with more cognitive failures tend to experience more errors (misjudgments and failures of observation) and lapses (absent-minded behaviors), and are involved in more accidents than those with fewer cognitive failures.

The results of this study produced other correlations of interest. Persons with more cognitive failures tend to experience more stress when driving than those with fewer cognitive failures. Persons with more cognitive failures are also more prone to be aggressive while driving. Finally, attributionally complex persons tend to be more alert while driving. These findings have implications for driver training as well as for increasing our understanding of the factors that influence driver error and safety.

Specifically, according to the results of this study, the "at risk" driver tends to experience more cognitive failures and to be less attributionally complex. As such, these persons pose a safety risk by committing more driving violations and by being likely to be aggressive and involved in more accidents. Persons with this combination of characteristics are likely to make more failures and misjudgments of observation, to commit absent-minded behaviors, to be less alert while driving, and to be anxious while driving. Further research investigating the cognitions of drivers is needed in order to determine the functionality of telematic use for these characteristically distracted "at risk" drivers.

References

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21*, 1-16.

Deffenbacher, J. L., Oetting, E. R., & Lynch, R. S. (1994). Development of a Driving Anger Scale. *Psychological Reports, 74*(1), 83-91.

Fletcher, G. O., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). Attributional complexity: An individual difference measure. *Journal of Personality and Social Psychology, 51* (4), 875-884.

Glendon, A. I., Dorn, L., Matthews, G., Gulian, E., Davies, D. R., & Debney, L. M. (1993). Reliability of the driving behaviour inventory. *Ergonomics, 36* (6), 719-726.

Herzog, T. A. (1994). Automobile driving as seen by the actor, the active observer, and the passive observer. *Journal of Applied Social Psychology, 24*, 2057- 2074.

Parker, D., Reason, J. T., Manstead, A. S. R., & Stradling, S. G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics, 38* (5), 1036-1048.

***This research was facilitated by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ARO under grant DAAD19-01-1-0621.**

110th Annual American Psychological Association Meeting, Chicago, August, 2002

Stress and Human Performance

Peter Hancock, Ph.D.

Department of Psychology and

Institute for Simulation and Training

University of Central Florida






UCF

INSTITUTE for SIMULATION & TRAINING

MURI

MIT²

On Laws and Laws of Behavior

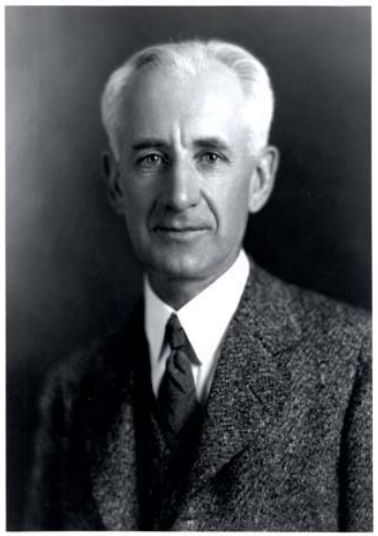
Laws—Description and Explanation

- It is Possible to Have Descriptive Laws
  - (e.g., Law of Inverse Squares, Newton, "Hypothesis non Fingo").
- It is Possible to Have Descriptive Laws that Intrinsically include Explanation
  - (e.g., Fitts Law – $MT = a + b\log_2 [2A/W]$)
- But the log base could be anything – it is only base two because of the link to information theory
- The truth is we don't like Laws without purported Cause.
- But worse are Laws that are Flawed.

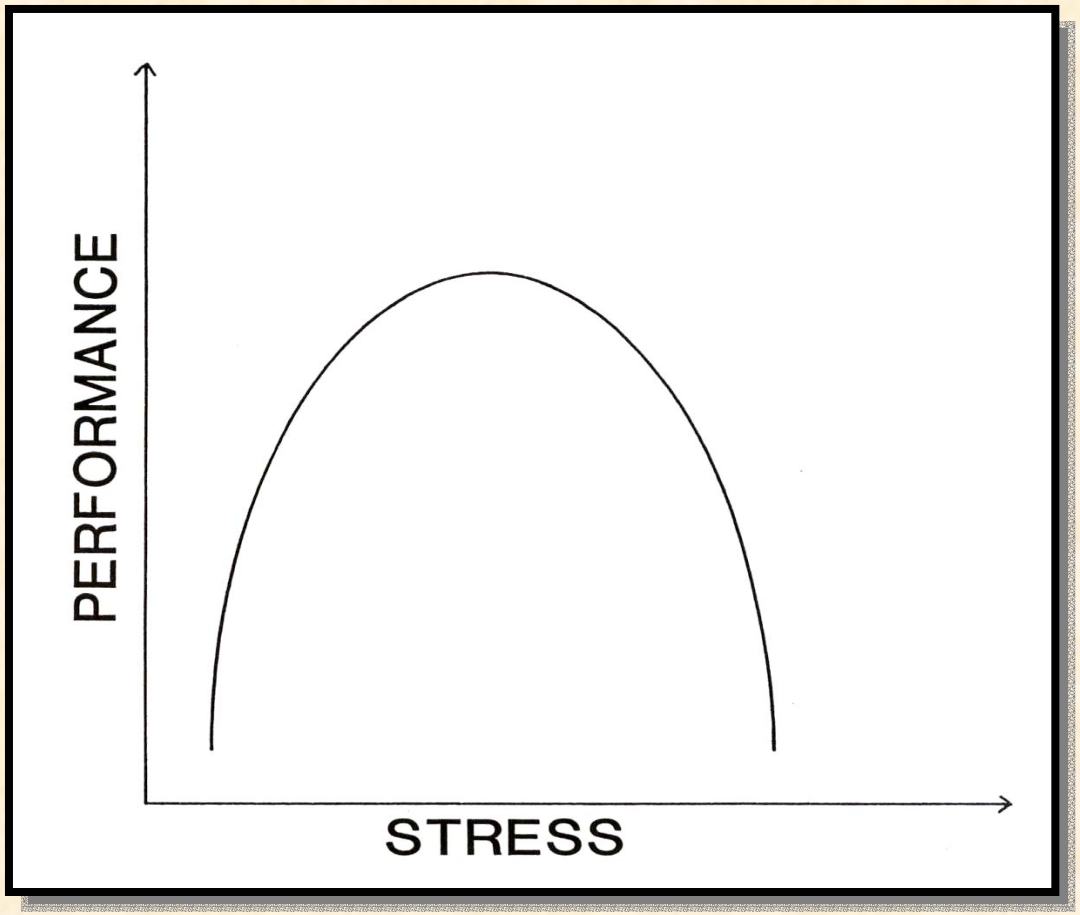
The Inverted-U

Stress and Human Performance

The Inverted-U Theory



Robert M. Yerkes





Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurological Psychology*, **18**, 459-482.

Stress and Human Performance

Textbook Citations

Table 1: *List of Introductory Psychology Texts That Give Reference To The Stress Performance Interaction as the Yerkes-Dodson Law.*

- Baron, R.A. (1998) *Psychology*, 4th Edition. Boston: Allyn & Bacon.
- Fernald, L.D., & Fernald, P.S. (1985) *Introduction to Psychology*, 5th Edition. Dubuque: Wm. C. Brown.
- Gergen, M.M., Suls, J.M., Rosnow, R.L., & Lana, R.E. (1989) *Psychology: A Beginning*. SanDiego: Harcourt, Brace, & Jovanovich.
- Haber, A., & Runyon, R.P. (1983) *Fundamentals of Psychology*. Reading: Addison-Wesley.
- Houston, J.P., Bee, H., & Rimm, D.C. (1983) *Invitation to Psychology*, 2nd Edition. New York: Academic Press.

Textbook Citations

Table 1: *List of Introductory Psychology Texts That Give Reference To The Stress Performance Interaction as the Yerkes-Dodson Law.*

- Lahey, B.B. (1983) *Psychology: An Introduction*. Dubuque: Wm. C. Brown.
- Lefton, L.A., & Valvatne, L. (1992) *Mastering Psychology*, 4th Edition. Boston: Allyn & Bacon.
- Morris, C.G., & Maisto, A.A. (2002) *Psychology: An Introduction*, 11th Edition. Englewood Cliffs: Prentice Hall.
- Roediger III, H.L., Rushton, J.P., Capaldi, E.D., & Paris, S.G. (1984) *Psychology*. Boston: Little, Brown, & Co.
- Seamon, J.G., & Kenrick, D.T. (1992) *Psychology*. Englewood Cliffs: Prentice Hall.
- Silverman, R.E. (1982) *Psychology*, 4th Edition. Englewood Cliffs: Prentice Hall.

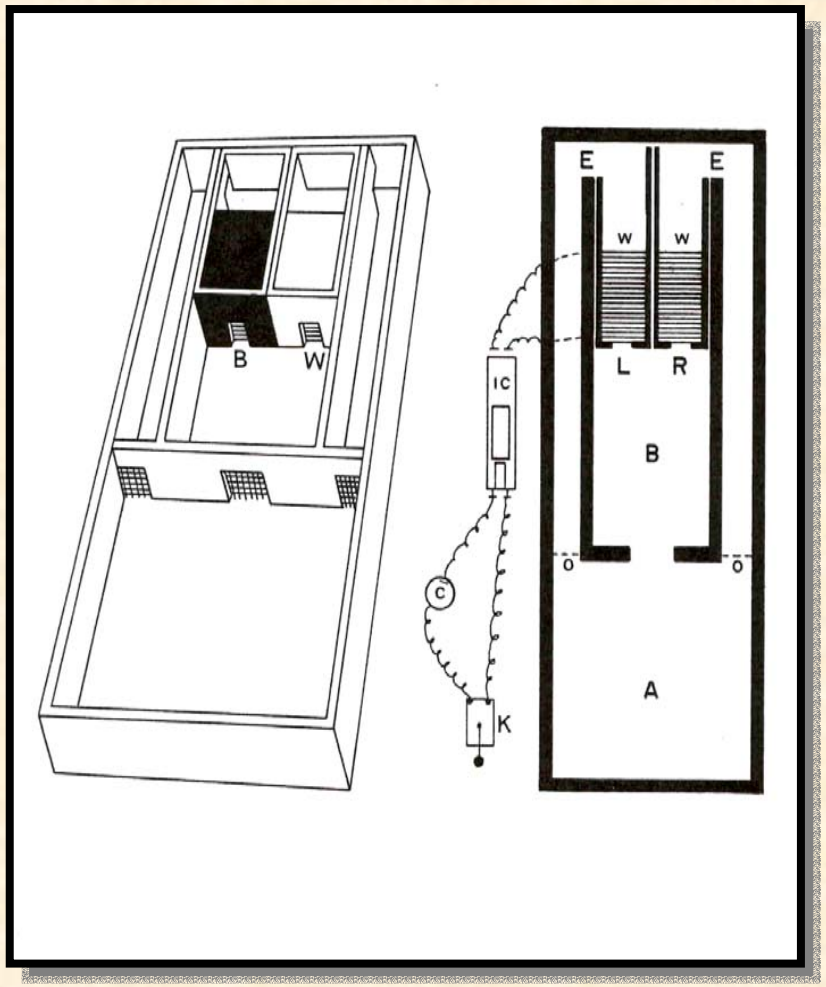
The Yerkes-Dodson Law

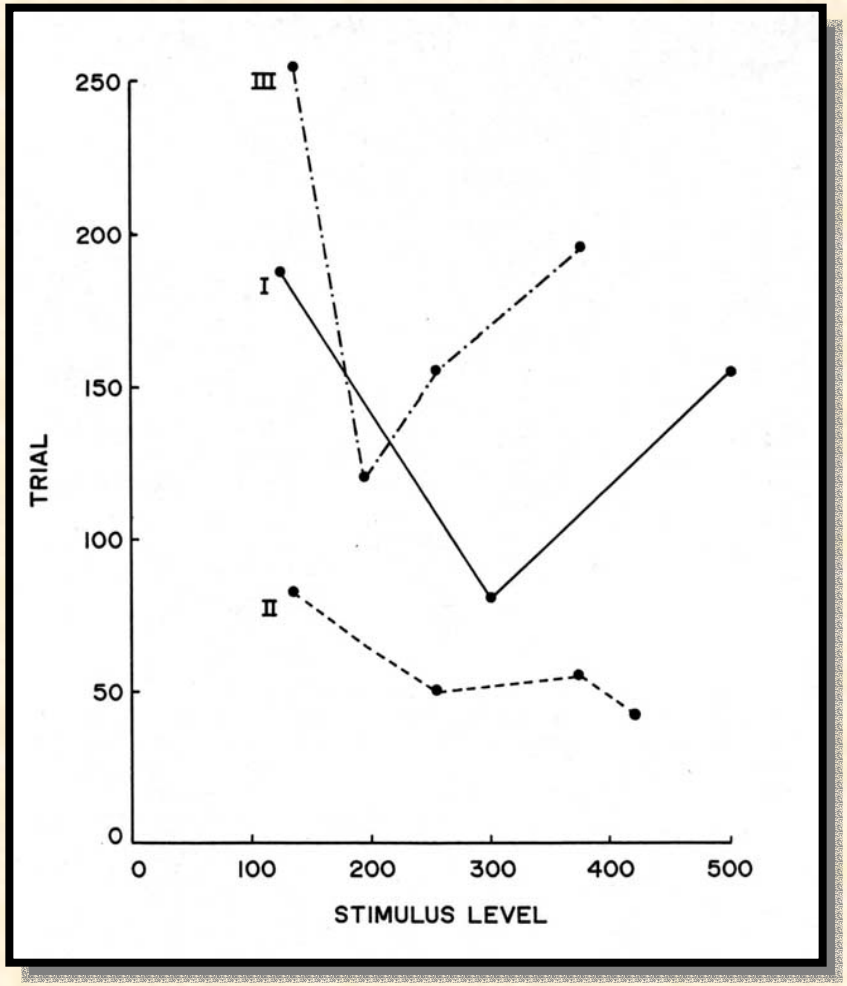
Rarely Recognized Facts

- Population: Of Mice not Men
- A 'Between-Mouse' Design
  - –Mus Musculus
  - –Known as dancing mice because they run around in little circles
  - –This behavior is due to a point mutation that causes deafness
- Some mice never learn
- Measured Learning not Performance
- Acute Electric Shock not Continuous Stress
- *The Critical result is based on Eight acutely shocked deaf, and genetically mutated mice*

Stress and Human Performance

The Original Experiment





Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurological Psychology*, **18**, 459-482.

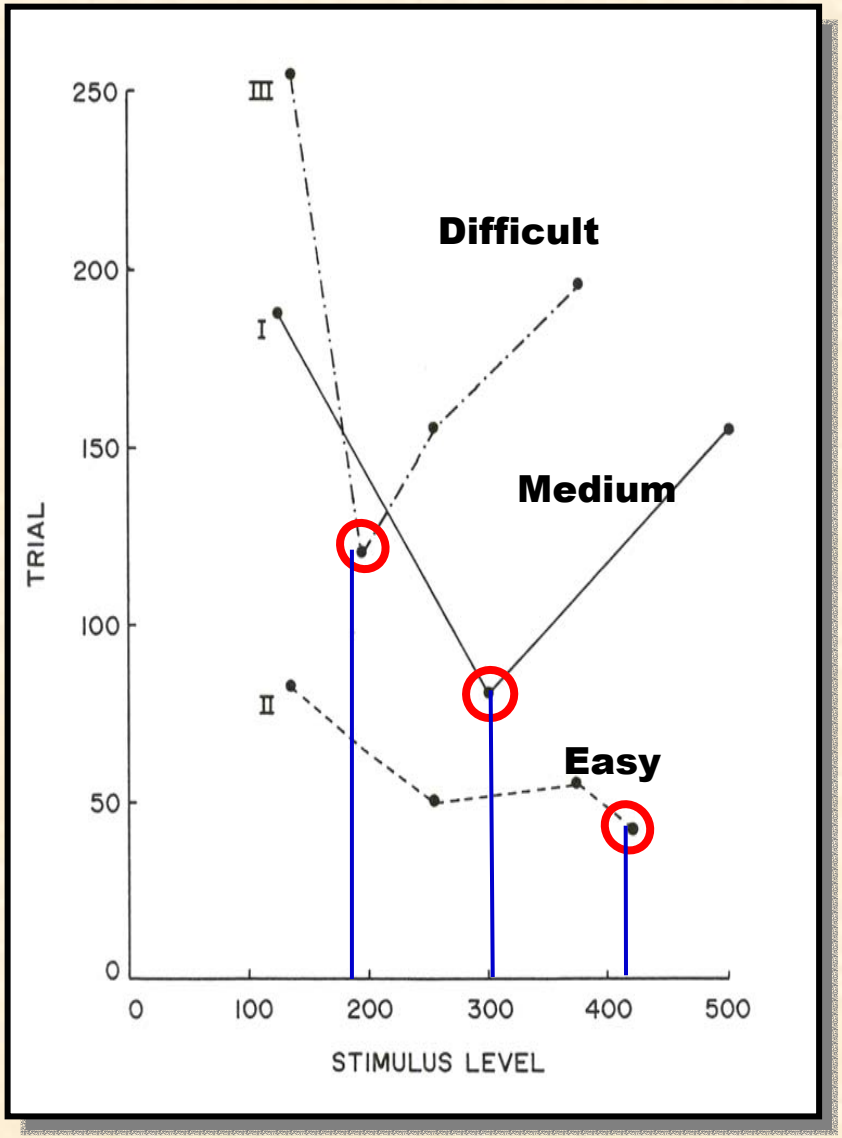
Stress and Human Performance

The Complexity Hypothesis

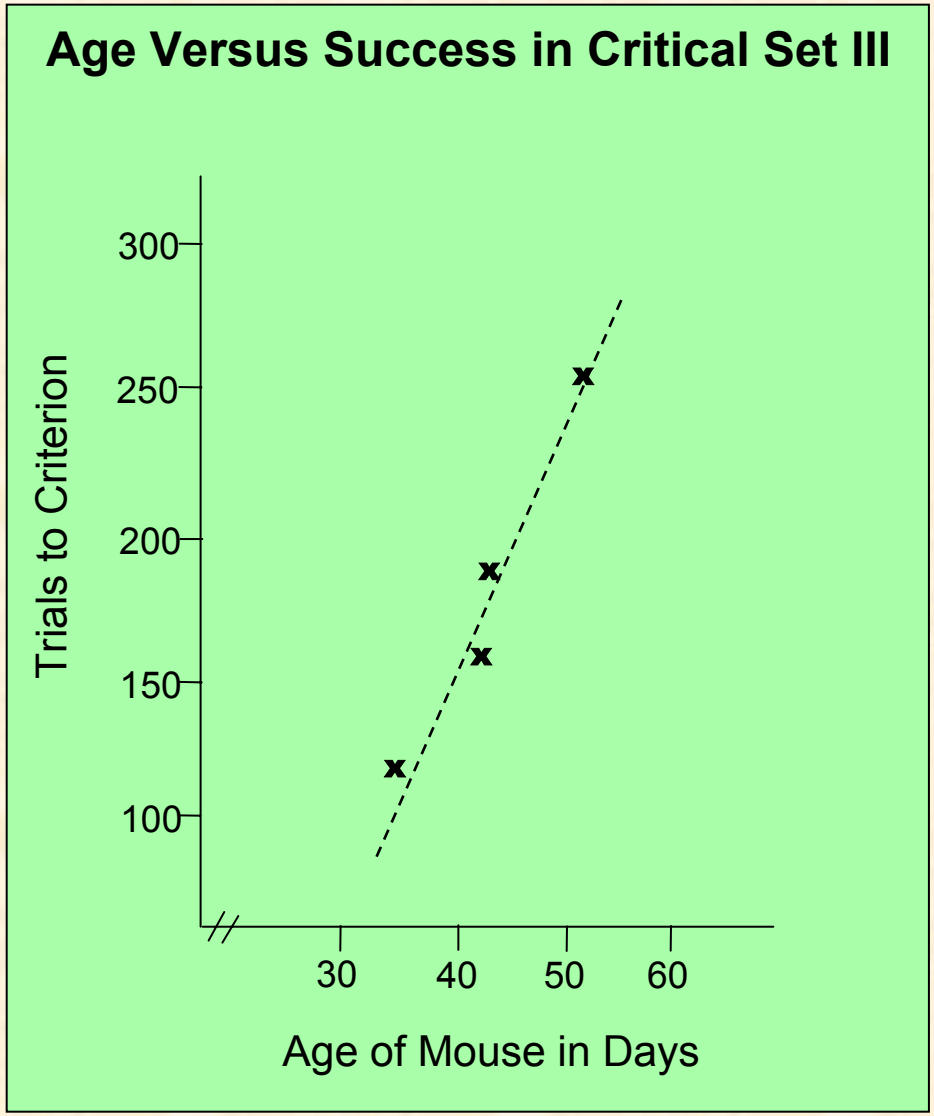
"As the difficulty of discrimination is increased the strength of that stimulus which is most favorable to habit-formation approaches threshold."

Yerkes and Dodson (1908, p. 481).

Thus, nominally the more complex the task, the less the level of stress to reach optimal performance.



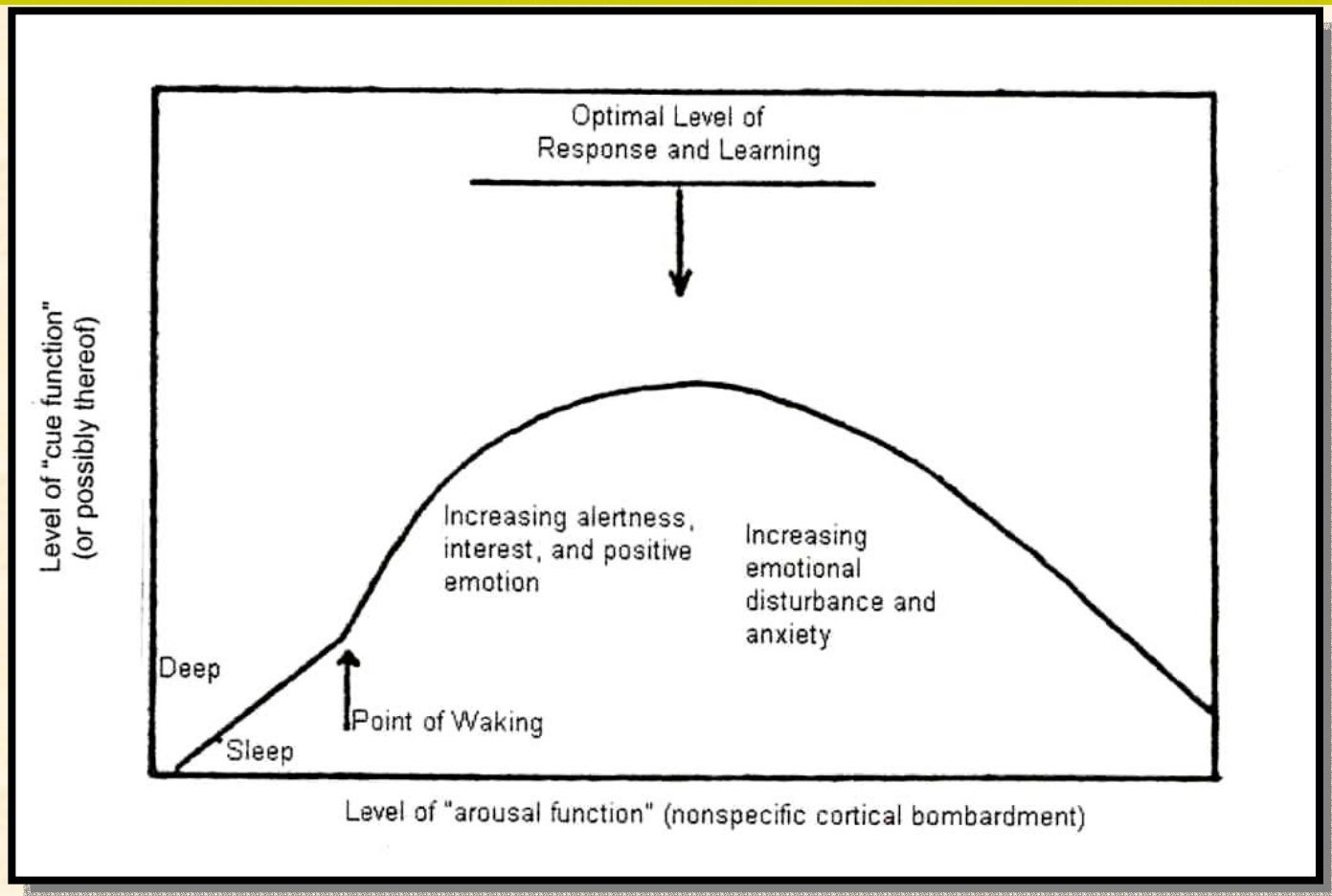
Problematic Confounds



Stress and Human Performance

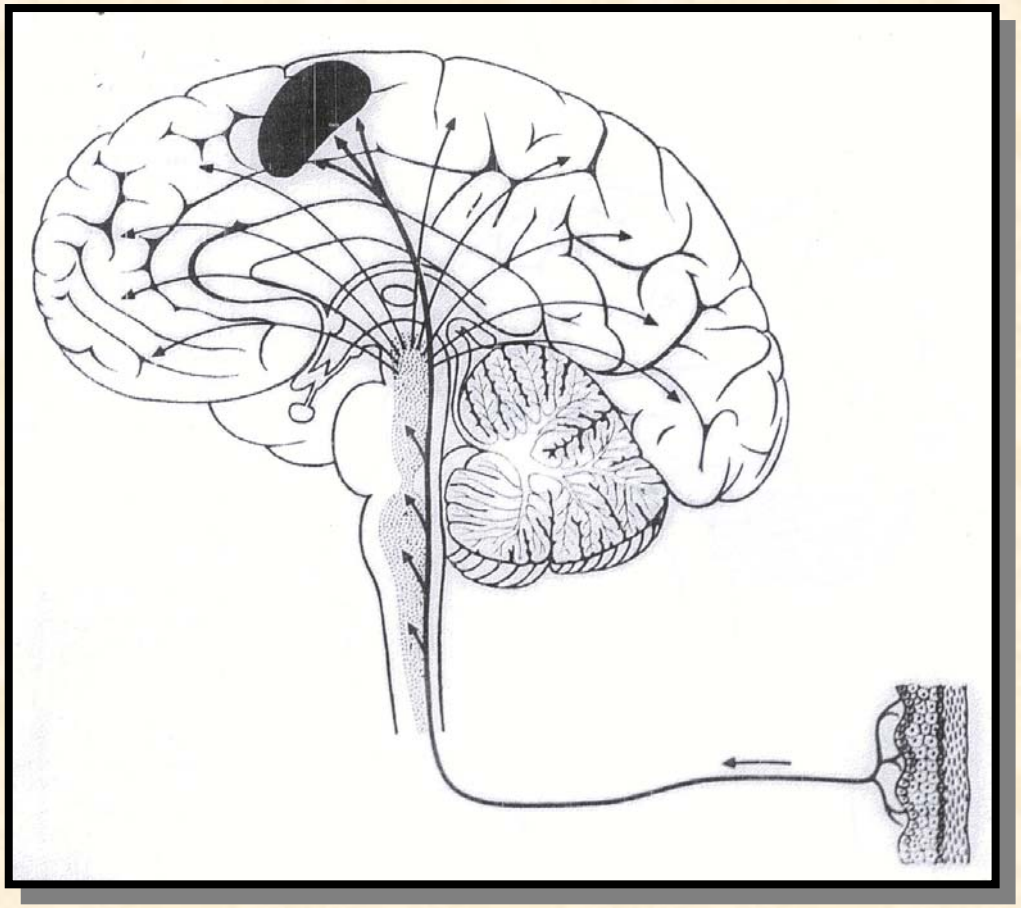
Resurrection of the Relationship

Hebb, D.O. (1955). Drives and the C.N.S. (conceptual nervous system). *Psychological Review*, **62**, 243-254.



Easterbrook, J.A. (1959). The effect of emotion on the utilization and the organization of behavior. *Psychological Review*, **66**, 183-201.

The Reticular Formation

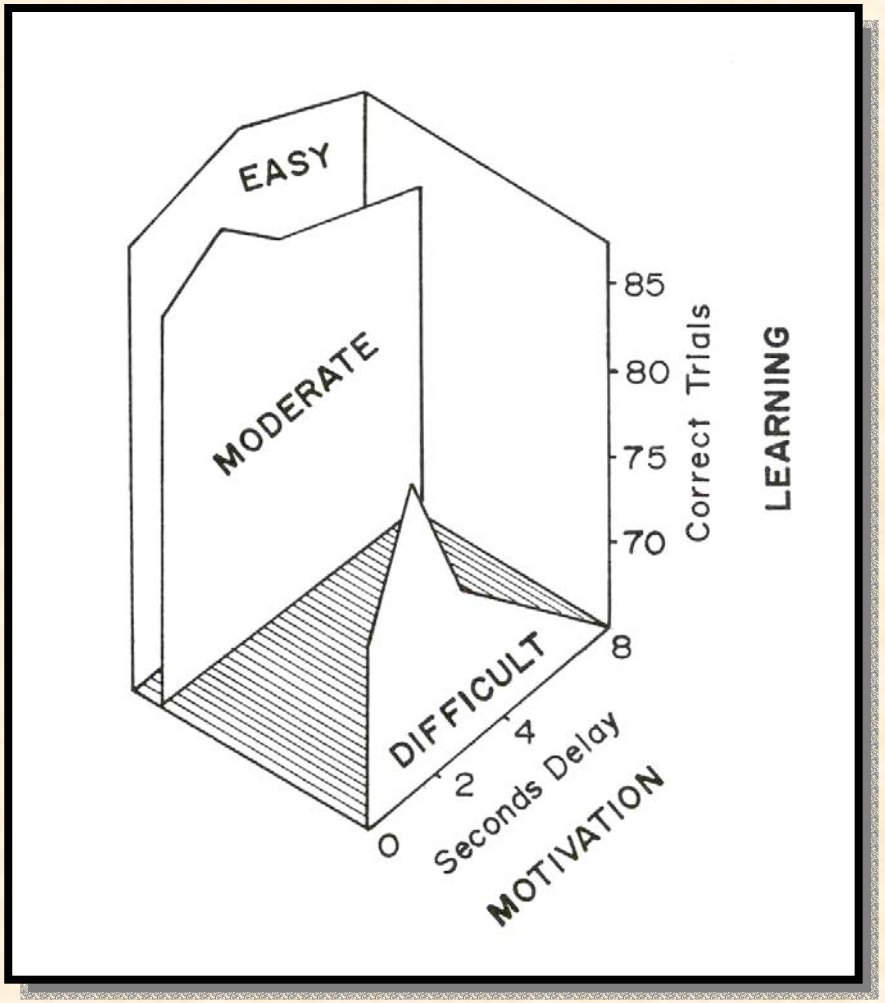


French, J.D. (1957). The reticular formation. *Scientific American*, **165**, (5) 54-60.
Moruzzi, G. & Magoun, H.H. (1949). Brain Stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, **1**, 455-473.

Empirical Support

Rats in a Water Maze





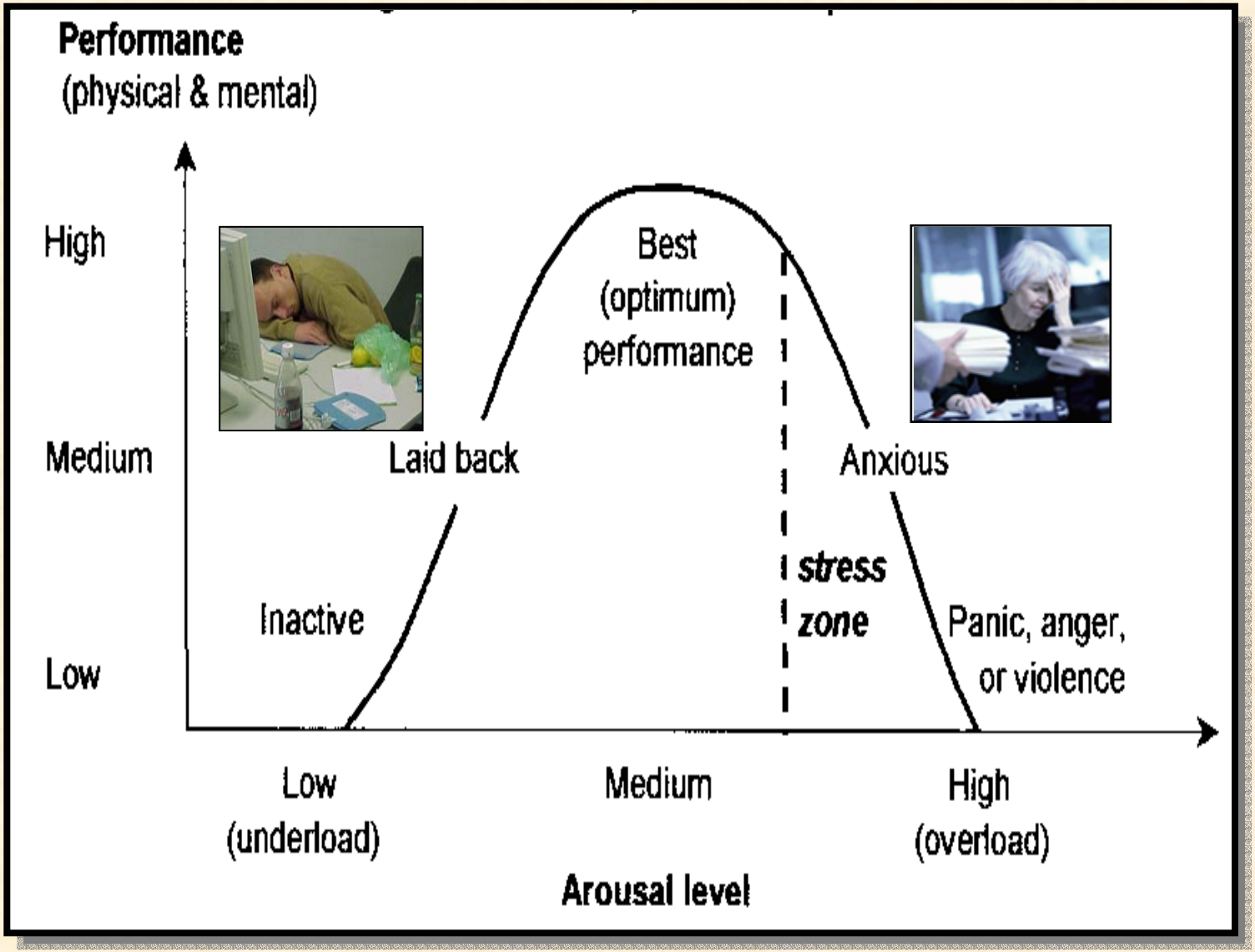
Broadhurst, P.L. (1957). Emotionality and the Yerkes-Dodson law. *Journal of Experimental Psychology*, **54**, 345-352.

Reasons for Longevity

- Just Plain Common Sense
- Never saw a data set it didn't like
- Arousal was an infinitely flexible 'explanation'

Stress and Human Performance

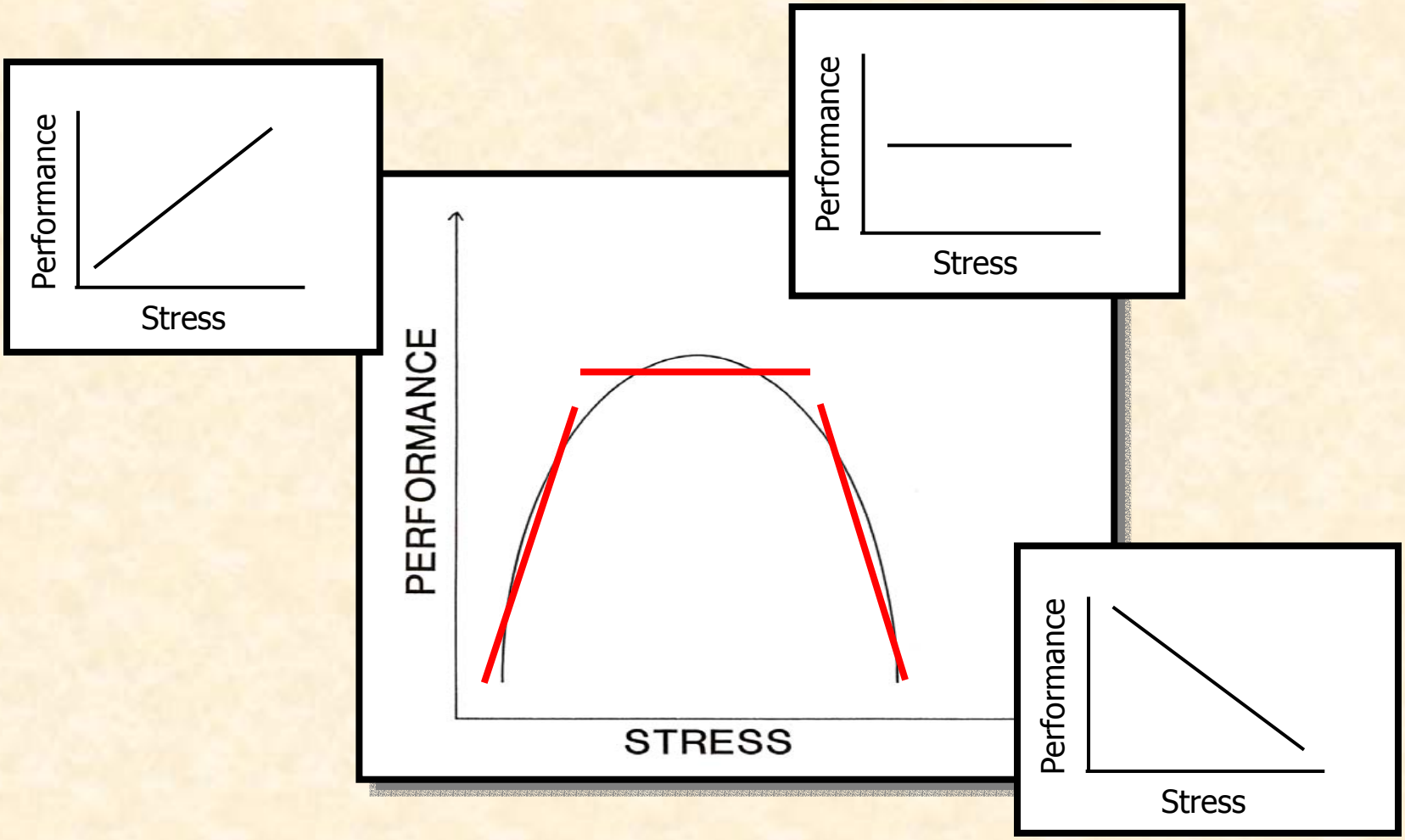
Just Common Sense



Stress and Human Performance

Post-Hoc Interpretation of Data

Virtually any pattern of results can be 'explained'



The Demise of Unitary Arousal

Stress and Human Performance

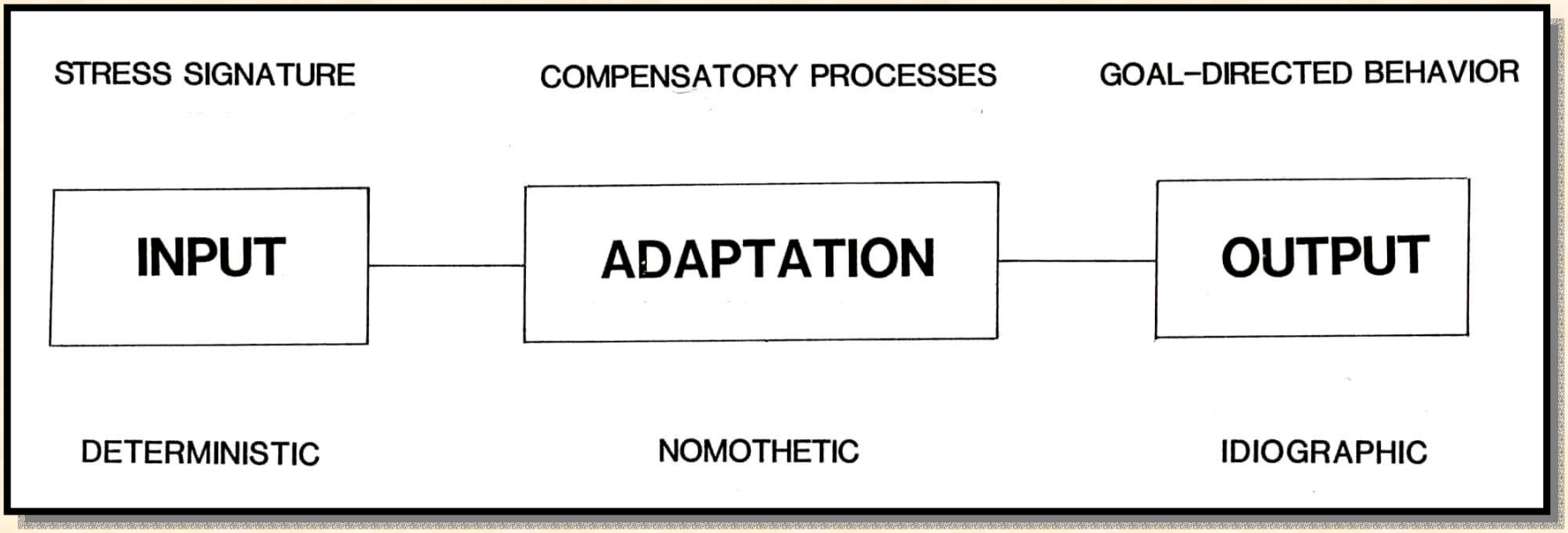
Patterning of Stress States

Hockey, G.R., & Hamilton, P.A. (1983). The cognitive patterning of stress states. In G.R. Hockey (Ed.), *Stress and fatigue in human performance,* 331-361. New York: Wiley.

Stressor or condition	Brief description of manipulation	Alertness	Selectivity	Fast responding Speed	Fast responding Accuracy	STM
Anxiety	High scorers on trait-state anxiety tests; induced by danger/threat	+	+	0	−	−
Noise	Continuous broad-band noise at about 80-100 dB	+	+	0	−	−
Amphetamine	Oral ingestion in normal clinical dosages	+	+	+	(?)	−
Nicotine	Smoking one or two cigarettes, or oral ingestion of nicotine tablets	+	+	+	0(?)	−
Incentives	Augmented feedback, financial reward, or induced competitiveness	+	+	+	+	+
Time of day (p.m.)	Testing at later times of day (between about 1500 and 2100)	+	(?)	+	−	−
Heat	WGBT of about 30-40°C for up to 2-3 hours	+(?)	+	0	−	0
Alcohol	Oral ingestion of about 0.2-0.5 g/kg (peak BA level about 0.04%)	−	+(?)	−	−	−
Minor tranquilizers	Oral ingestion of benzodiazapines in normal clinical dosages	−	−	−	−	−
Sleep loss / night-work	One night without sleep/working at night before adaptation	−	−	−	−	0
Prolonged work / boredom	Continuous work for 30+ minutes on repetitive tasks	−	+(?)	−	−	0
Time of day (a.m.)	Testing at earlier times of day (between	−	(?)	−	+	+

The Maximal Adaptability Model

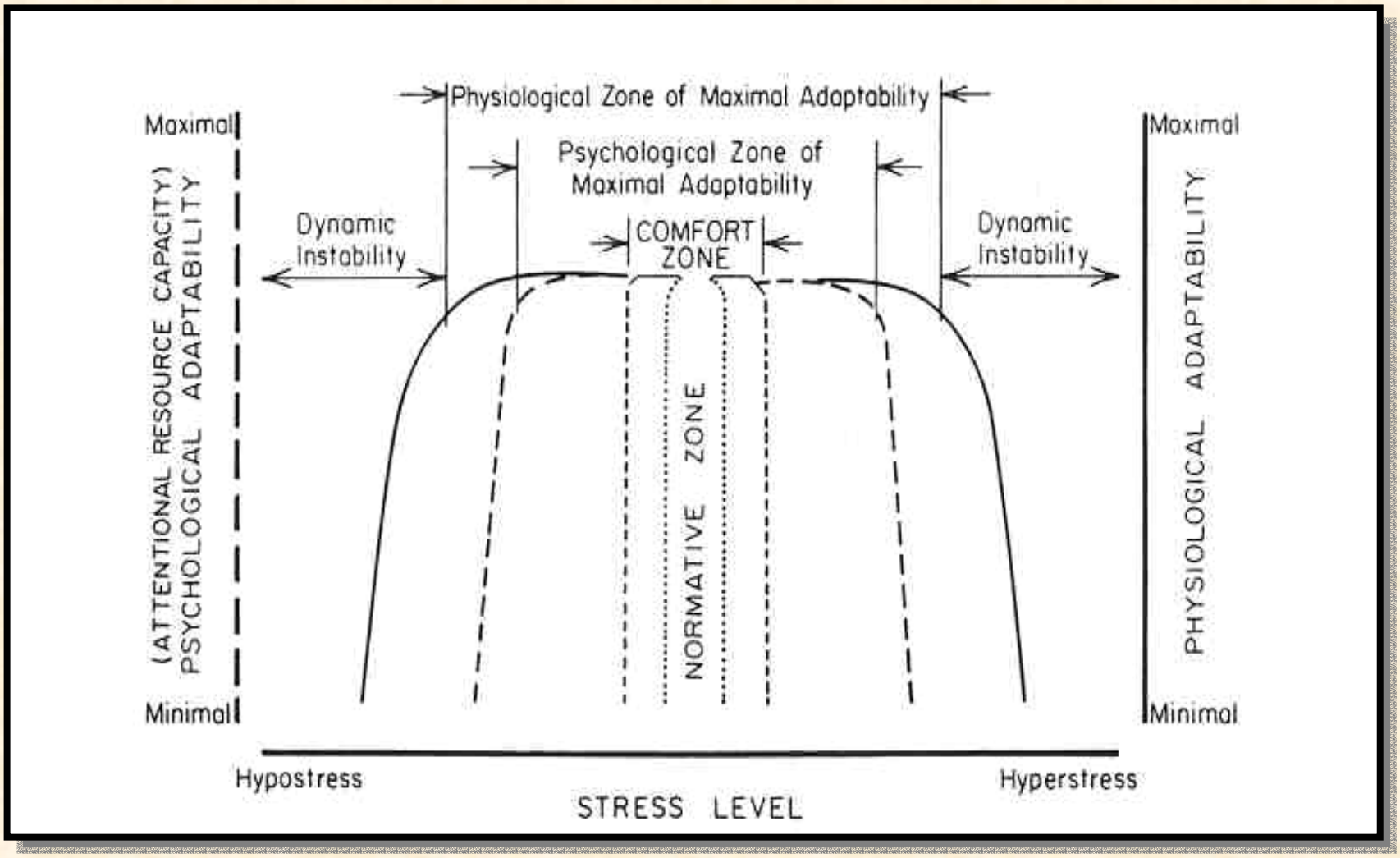
The Trinity of Stress



Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, **31**, 519-537.

Stress and Human Performance

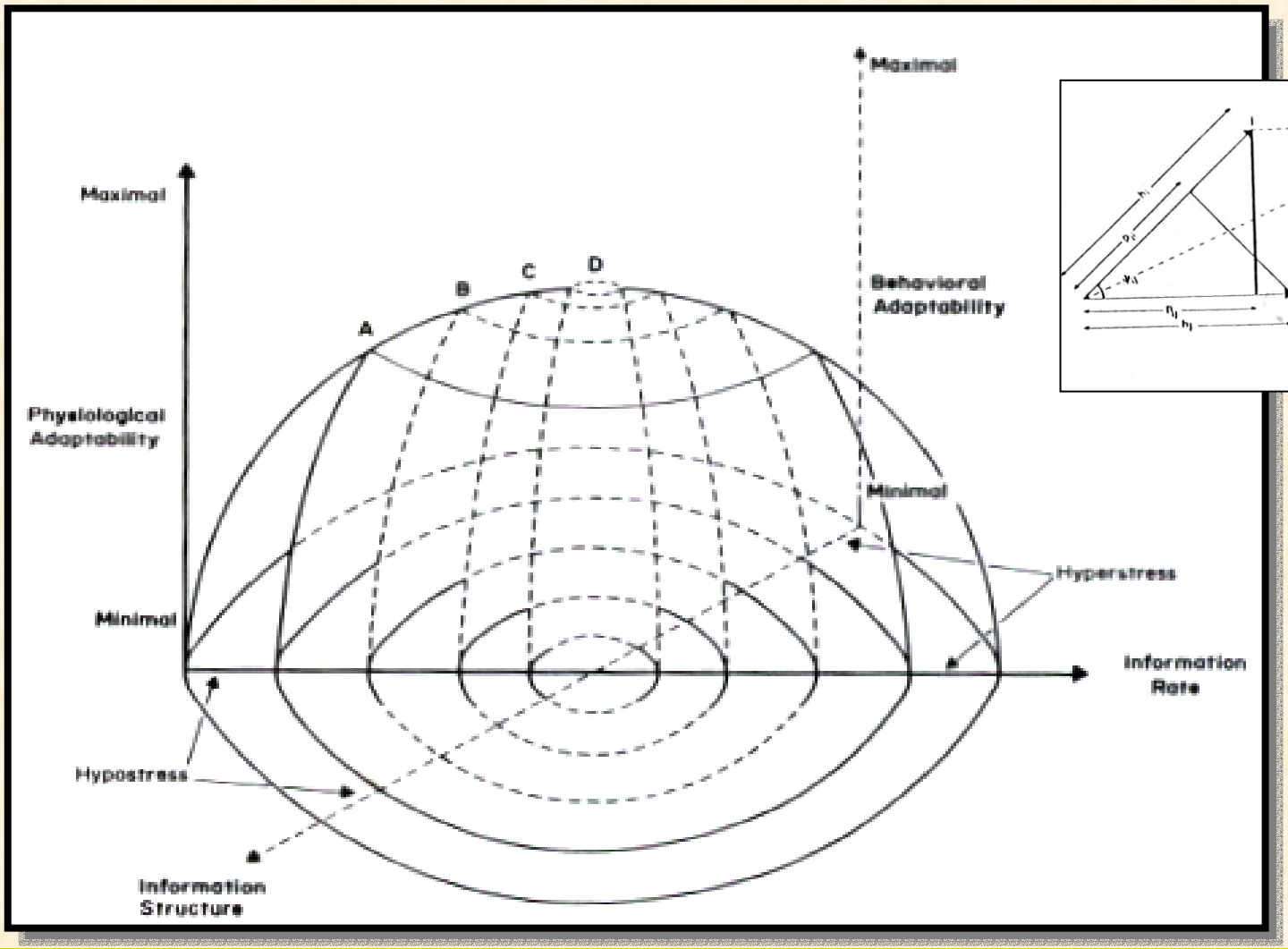
The Extended - U

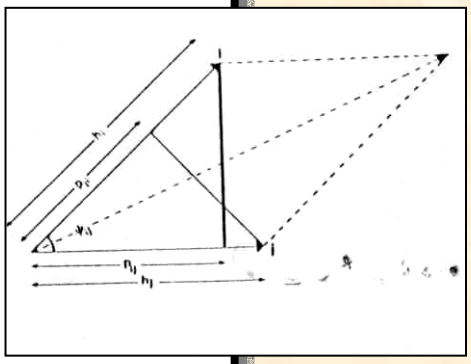


Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, **31**, 519-537.

Stress and Human Performance

Integrating Multiple Sources of Stress

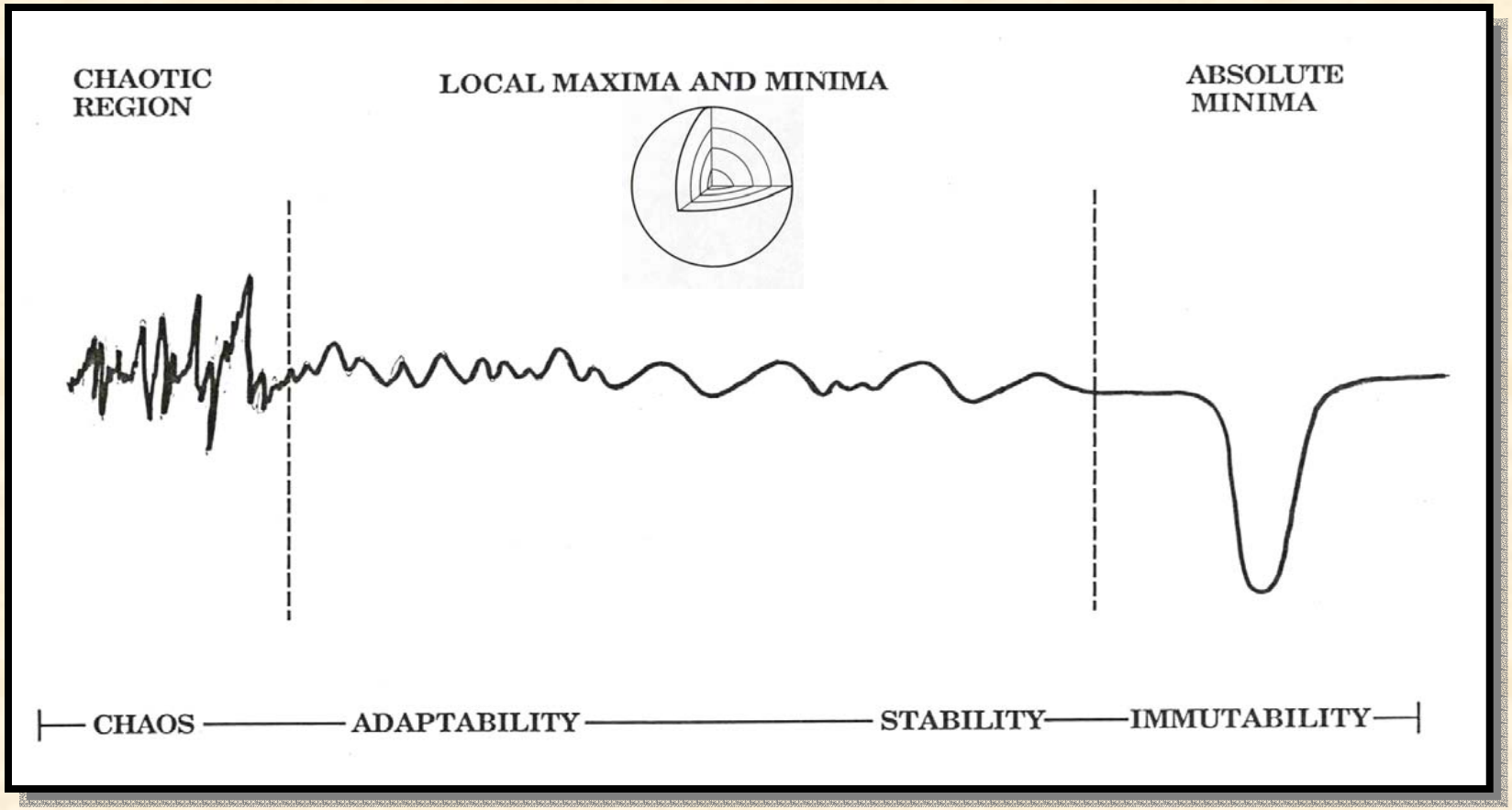




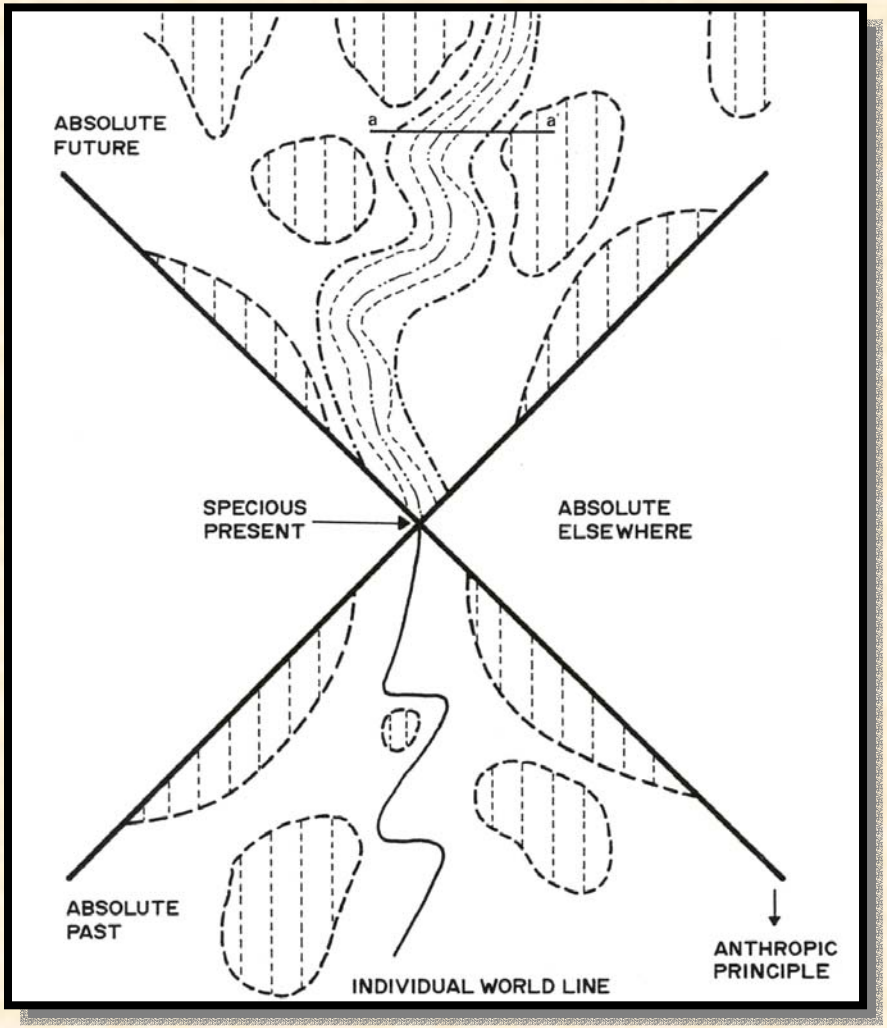
Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, **31**, 519-537.

Stress and Human Performance

Between Chaos and Immobility

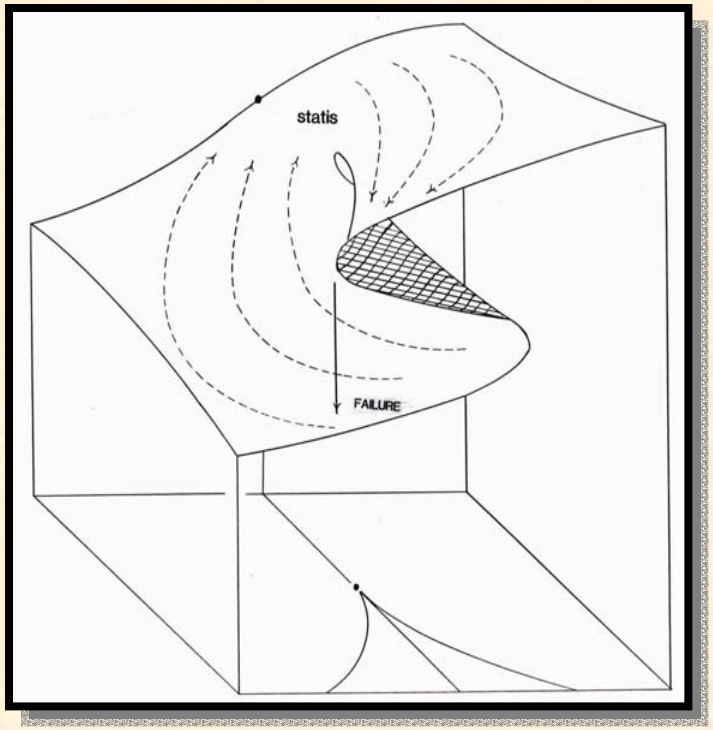


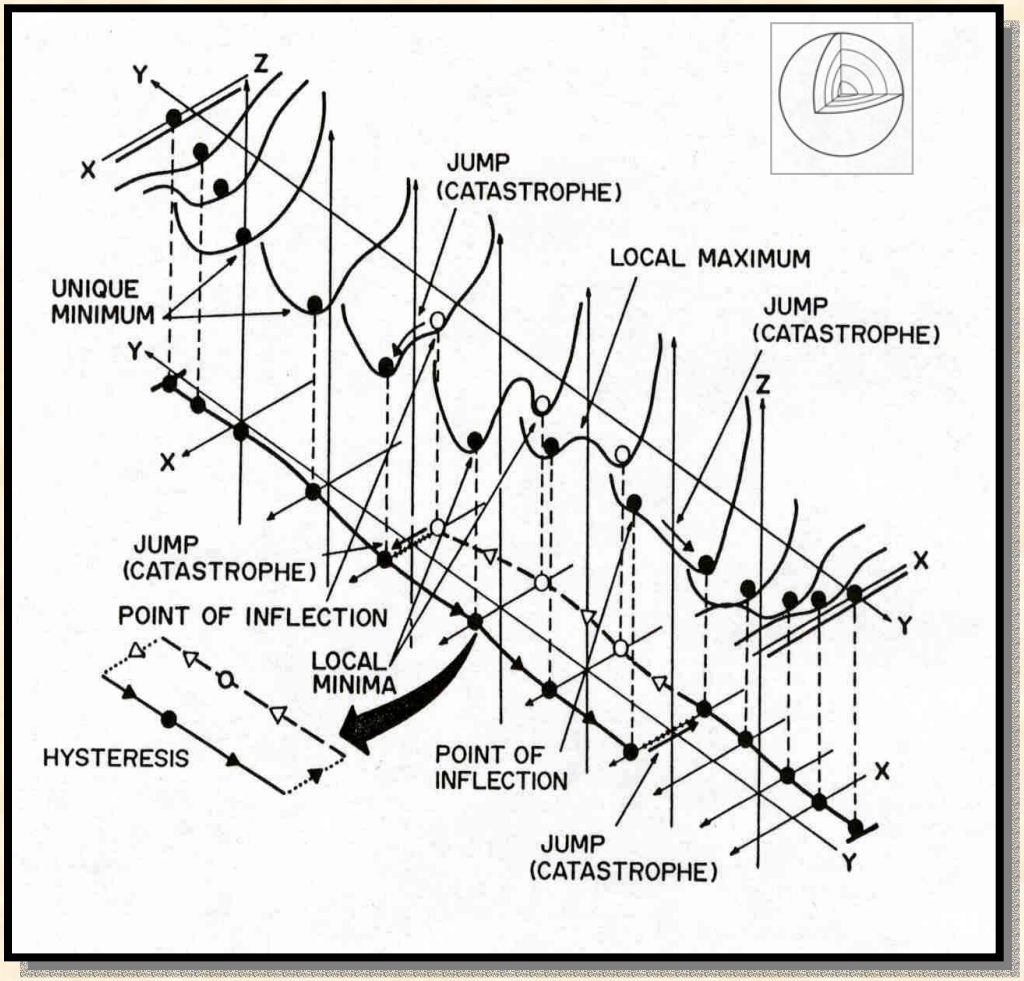
Spatio-Temporal Progress



Stress and Human Performance

Spatio-Temporal Progress

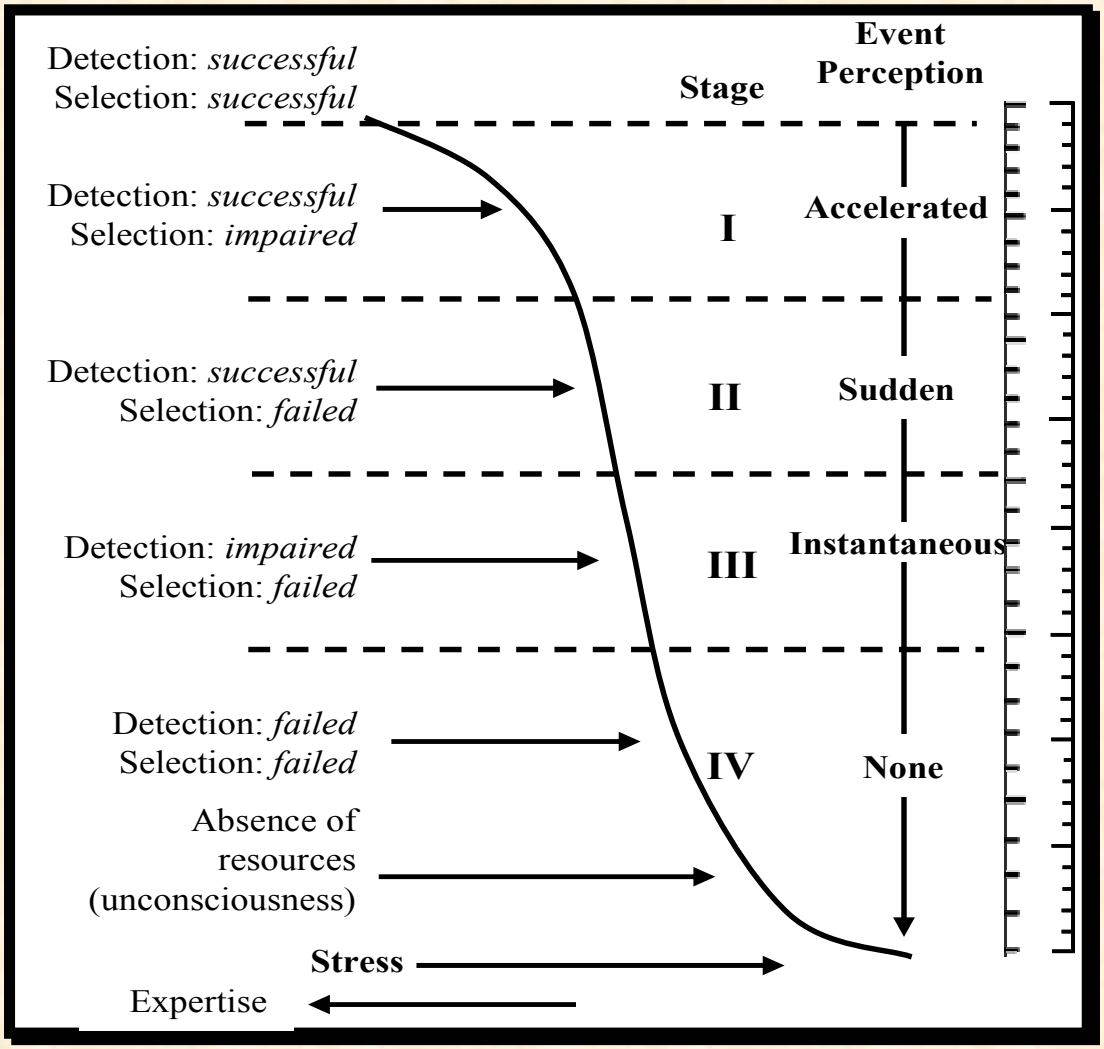


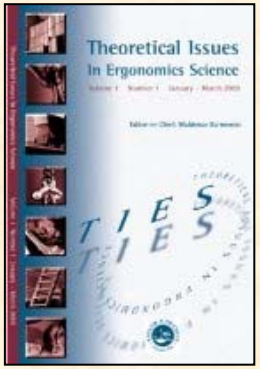


Kugler, P.N., & Turvey, M.T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. L. Erlbaum Associates: Hillsdale, N.J.

Stress and Human Performance

Performance at the Very Limit of Capability

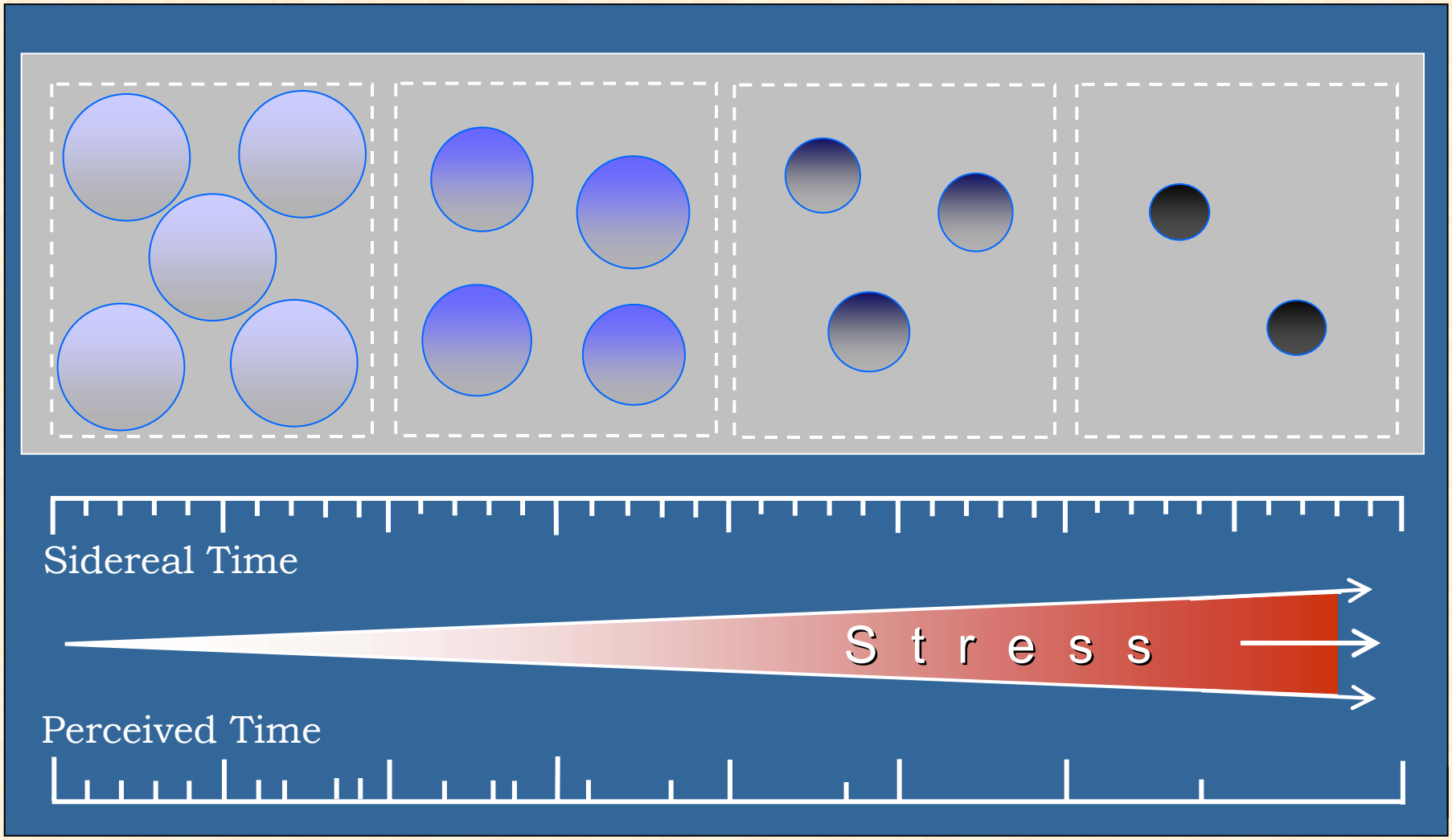




Hancock, P.A., & Weaver, J.L. (2002). Temporal distortions under extreme stress. *Theoretical Issues in Ergonomic Science*, in press.

Operator Performance Under Stress

Load-Shedding & Attentional Strategy



Stress and Human Performance

MURI Contributing Researchers



Peter Hancock, Ph.D.
Department of Psychology
University of Central Florida, IST



Mustafa Mouloua, Ph.D.
Department of Psychology,



Jim Szalma, Ph.D.
Department of Psychology



Clint Bowers, Ph.D.
Department of Psychology, TPL,



Jeanne Weaver, Ph.D.
Department of Psychology



Richard Gilson, Ph.D.
Department of Psychology, CAHFA



Eduardo Salas, Ph.D.
Department of Psychology



Mark E. Koltko-Rivera, Ph.D.
Department of Psychology



Shawn Burke, Ph.D.
Department of Psychology, IST



Raja Parasuraman
Cognitive Science Laboratory
Catholic University



Arthur S. Leon, M.D.
Laboratory of Physiological Hygiene and Exercise Science
University of Minnesota.



Wayne Harris, Ph.D.
Human Factors Research Laboratory
University of Minnesota



Kip Smith, Ph.D.
Department of Psychology Kansas State University



Ava Walker, Ph.D.
Department of Kinesiology
University of Minnesota

Stress and Human Performance

Contributing Researchers

UCF Human Factors Psychology Graduate Students


Christina Morris


Adams Greenwood-Ericksen


Krystal Bradley – Radakovich


H.C.Neil Ganey


Jennifer Thropp


Derek D. Diaz


Jessica Helmick


Chris Brill


Lorrenzo Torrez

Undergraduate & Post Bac. Research Assistants


Jennifer Ross


Danielle Adzima


Shawn Stafford


Chrissy Giddens


Craig Russell

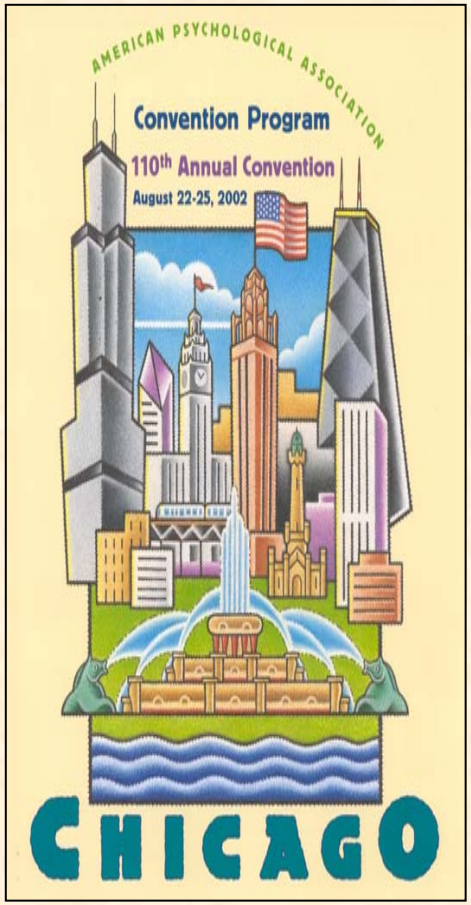

Cristina Vega

Coordinator Administrative Services


Barbara Koehler
Coordinator, Administrative Services, UCF

Stress and Human Performance

Acknowledgement



The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. This work was also made possible by support from the Advanced Decision Architectures Collaborative Technology Alliance Sponsored by the US Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009.

Our thanks to Elmar Schmeisser, Sherry Tove , Mike Dillings, Rene De Pontbriand, Laurel Allender, and Lou Bandaret



OPUS

Operator Performance Under Stress



University of Central Florida

Kansas State University

Catholic University

University of Minnesota

MURI

MIT

INSTITUTE for SIMULATION & TRAINING

THE CATHOLIC UNIVERSITY OF AMERICA

Multi-Disciplinary University Research Initiative

MIT²

THE CATHOLIC UNIVERSITY OF AMERICA

INSTITUTE for SIMULATION & TRAINING

Operator Performance Under Stress

MURI

MIT²

The Catholic University Of America

Institute for Simulation & Training

University of Central Florida	Catholic University	University of Minnesota	Kansas State University
MIT²	**IST**		
Time Distortion	Teams Under Stress	Brain States & Stress	Physical & Mental Workload	Decision-Making
Interfaces for Stress	Stress & Team Cognition	Cognitive Workload	Physical Evaluation	Distrust
Visual Search	Low Cost Training			
Delta Force	Flashpoint			
Viking		Viking	Viking	Biomedicine

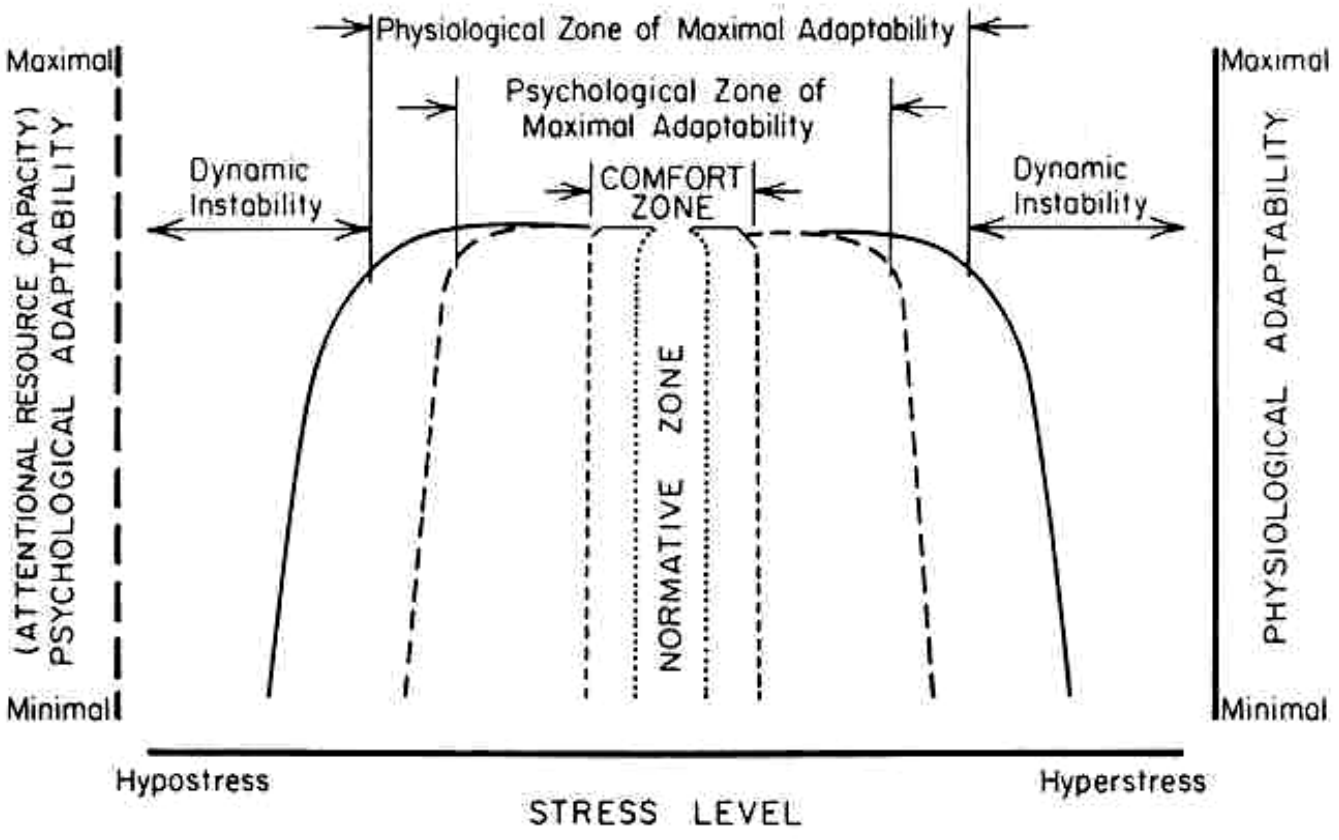
New Affiliated Universities

University of Cincinnati
(Stress Transitions – High/Low Workload)

University of Toronto
(Wearable Computers for Combat)

Florida Institute of Technology
(Adaptive Interfaces for Stress)

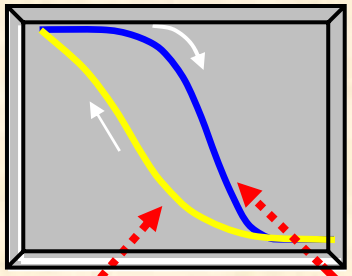
Stress and Performance Theory

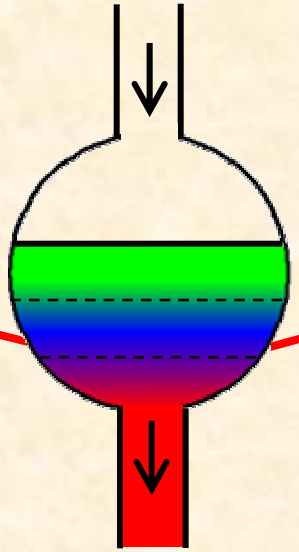


Hancock, P.A., & Warm, J.S., (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.

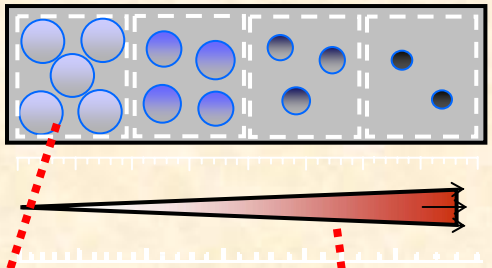
Attentional Depletion/Repletion

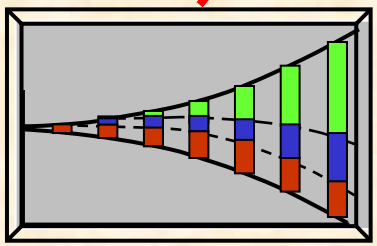
Hystersis
Failure and Recovery



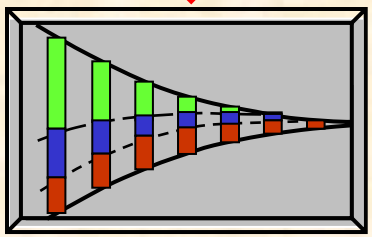


Stress
Distorted Perceptual Space-Time

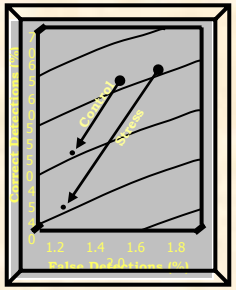




SKR Recovery



Stress Depletion

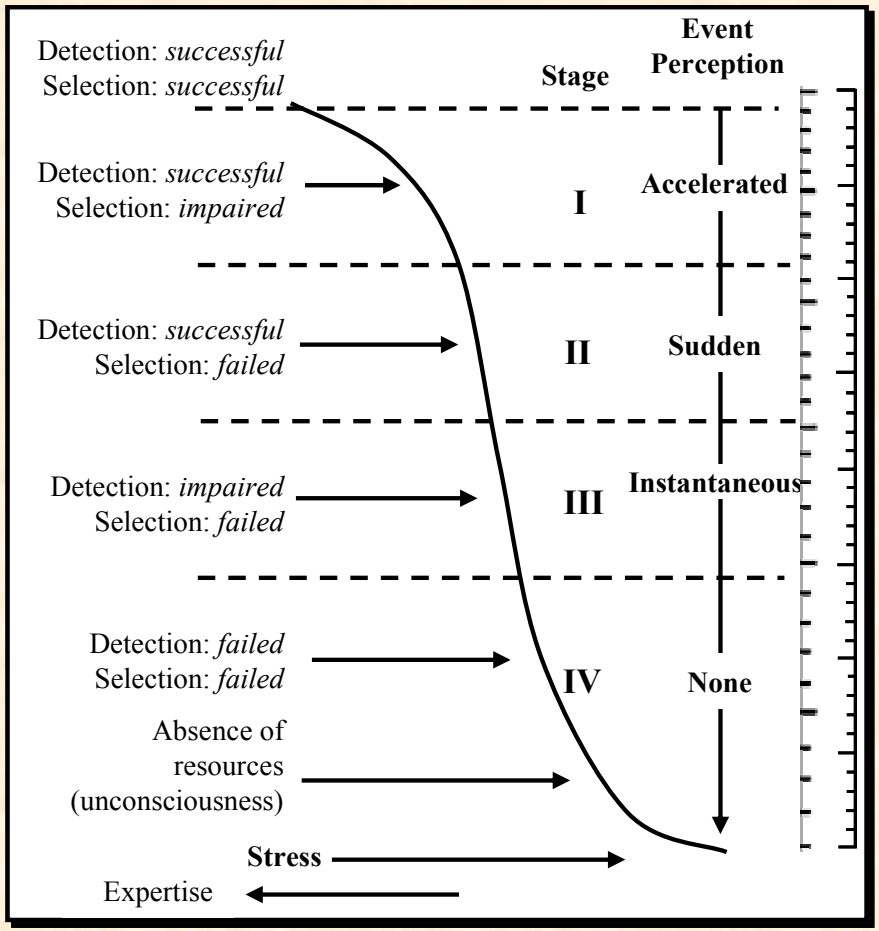


Stress and SDT



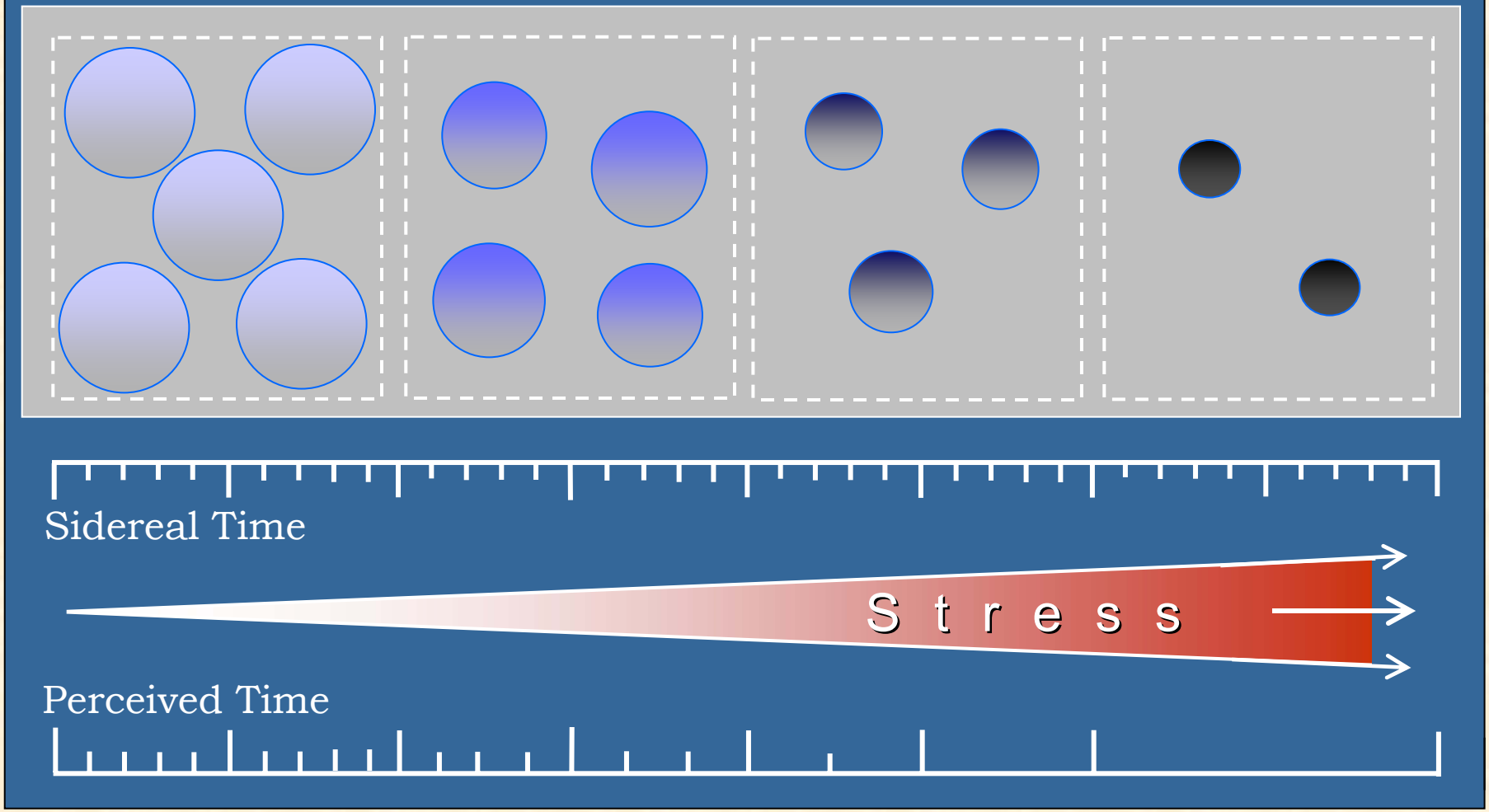
Time Distortion

Stress and Performance Theory



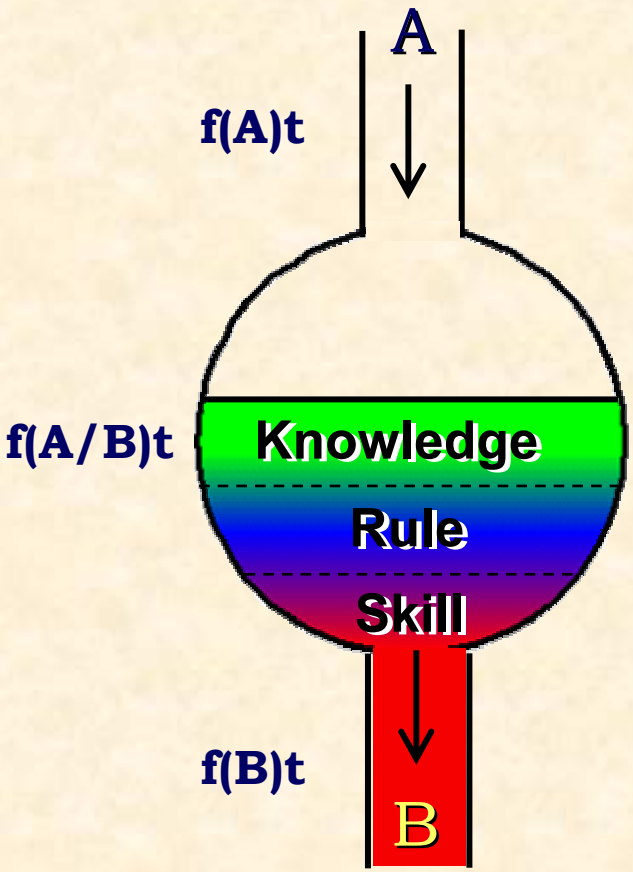
Hancock, P.A., & Weaver, J.L. (2001). Temporal distortions under extreme stress. Theoretical Issues in Ergonomics, VOL.

Stress Distorts Perceptual Space/Time



Stress and Attention

Resource Depletion or Resource Replacement?

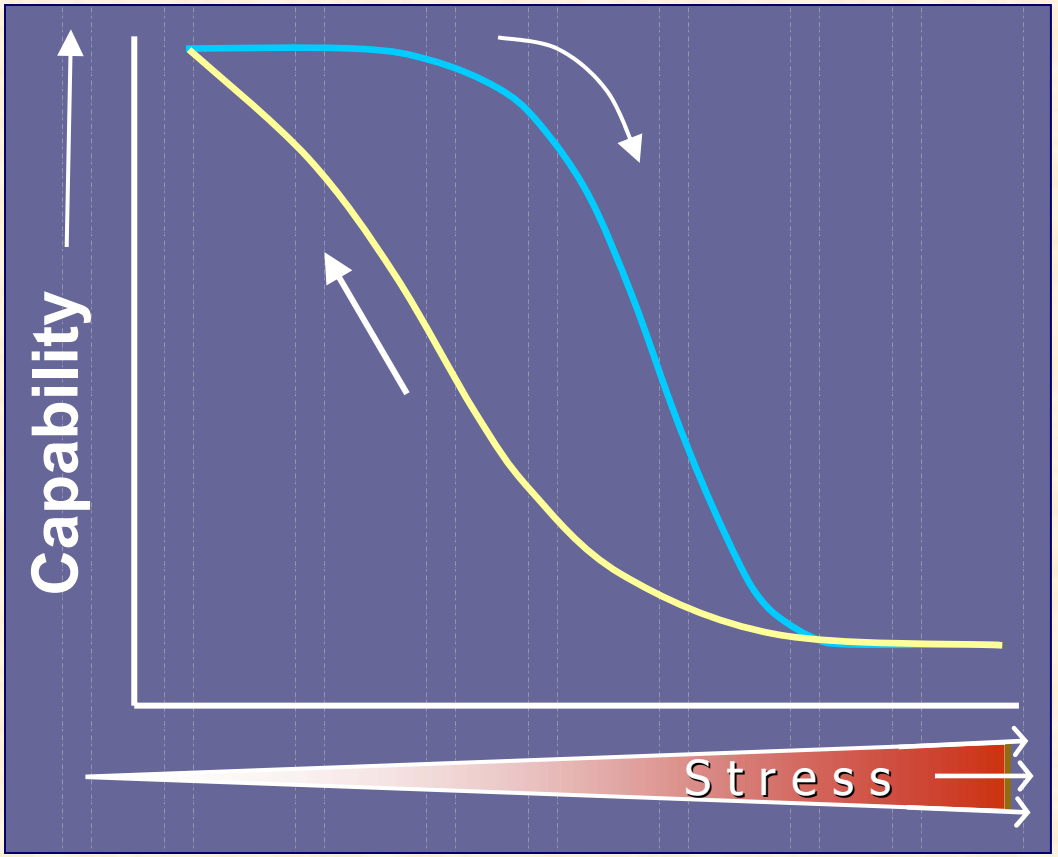


Unexplored aspects of resource change include:

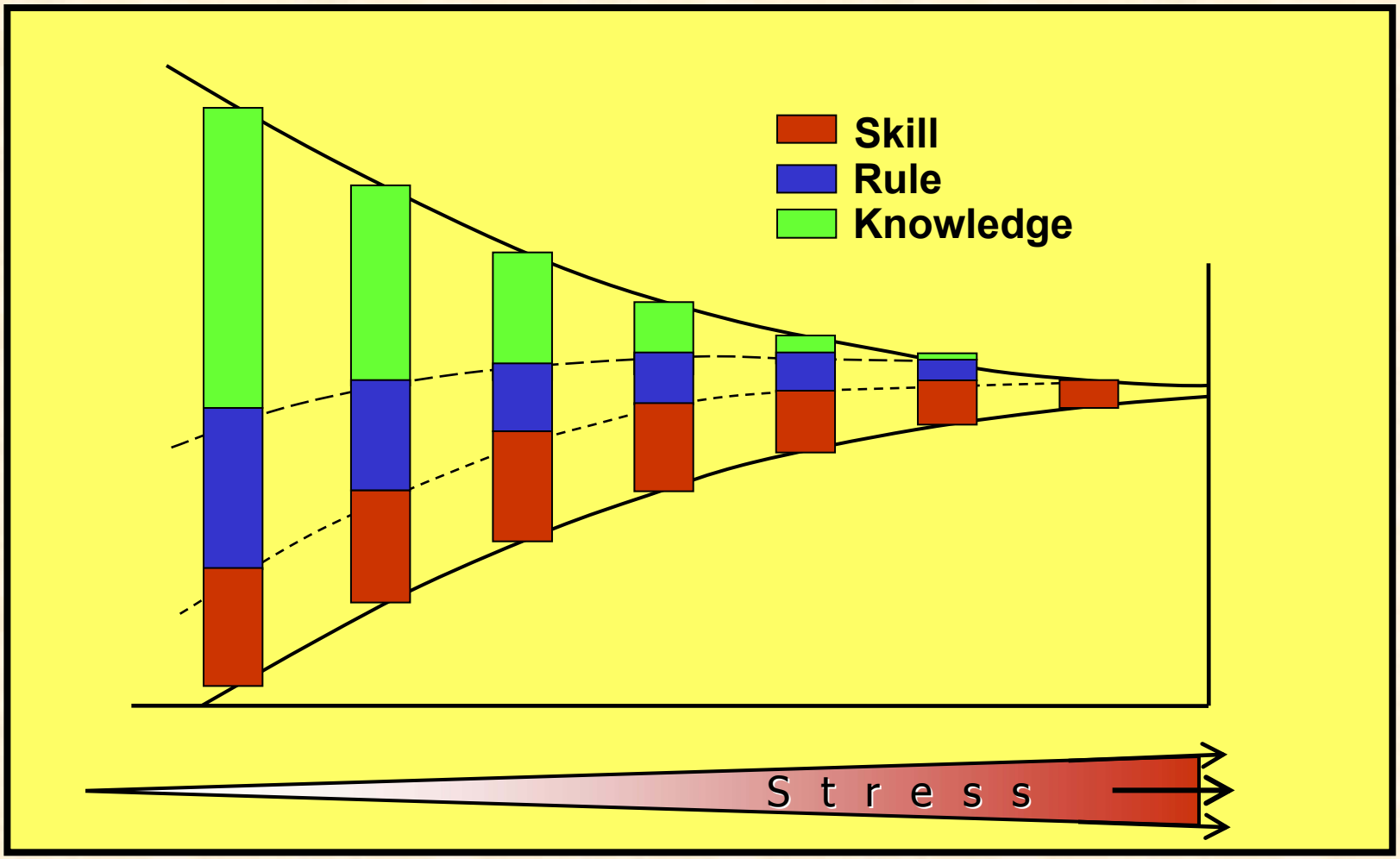
1) Depletion vs. Replenishment
2) Differential Resource Quality
3) Thresholds for Resource Depletion
4) Hystersis in Loss and Recovery

Hysteresis: Stress-Induced Failure and Recovery

Is the way in which the Individual recovers symmetrical with the way the individual fails?



Load-Shedding and Attentional Strategy

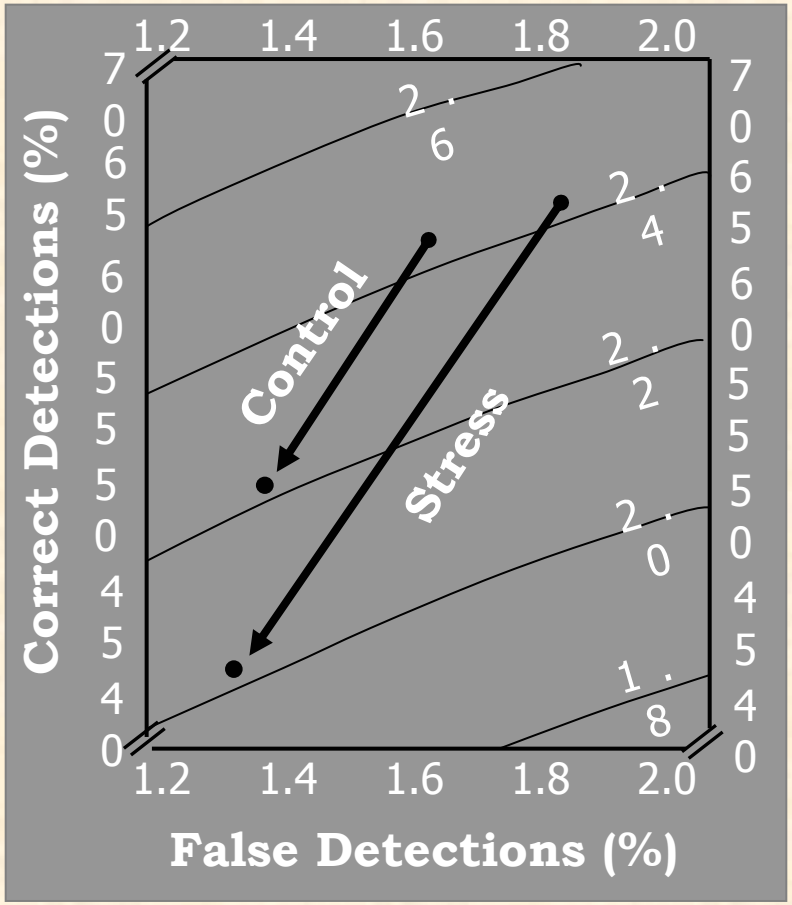


Stress Recovery



Stress and SDT

- How Does Stress Impact d' and B?
- Stress Reduces Positive Responses?
- Fuzzy Signal Detection under Stress?



Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, 42, 636-659.

Is Search Memory Disrupted By Noise Stress?

"Search memory"

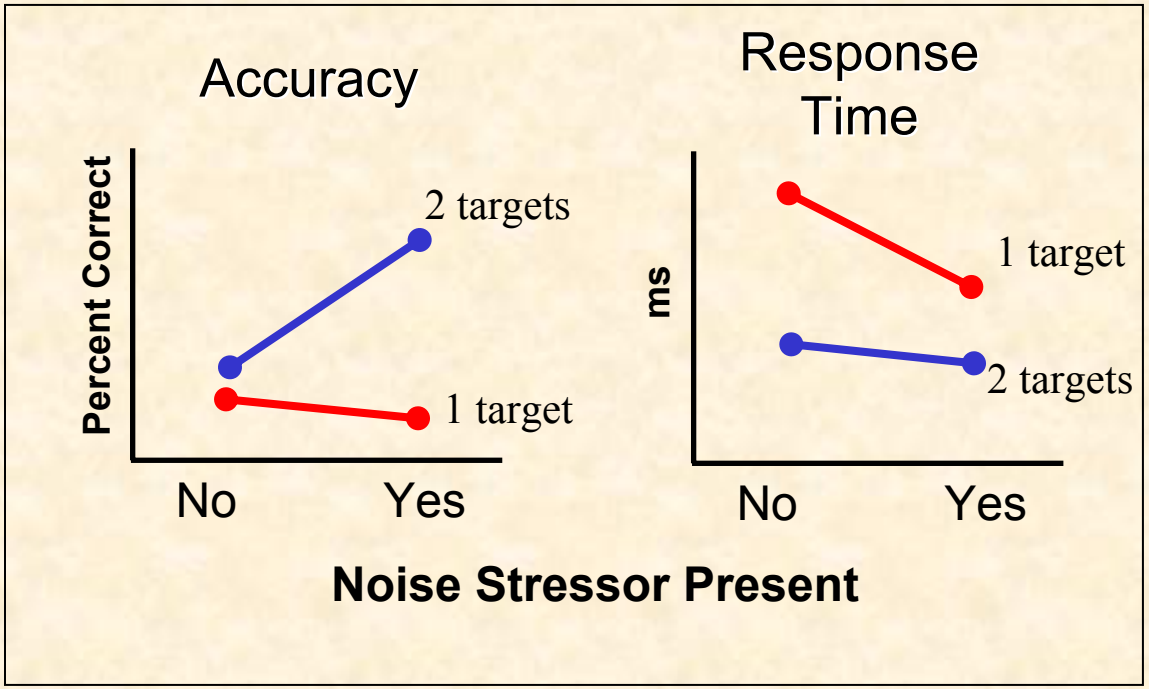
- Memory for items that have already been searched
- Facilitates efficient search
- Guides visual attention to new objects



D.D. Diaz, P.A. Hancock, & V.S. Sims. (2002). Visual Search for Two Identical Targets: The Role of Ecological Validity. Annual APA Midyear Symposium. Ft. Belvoir.

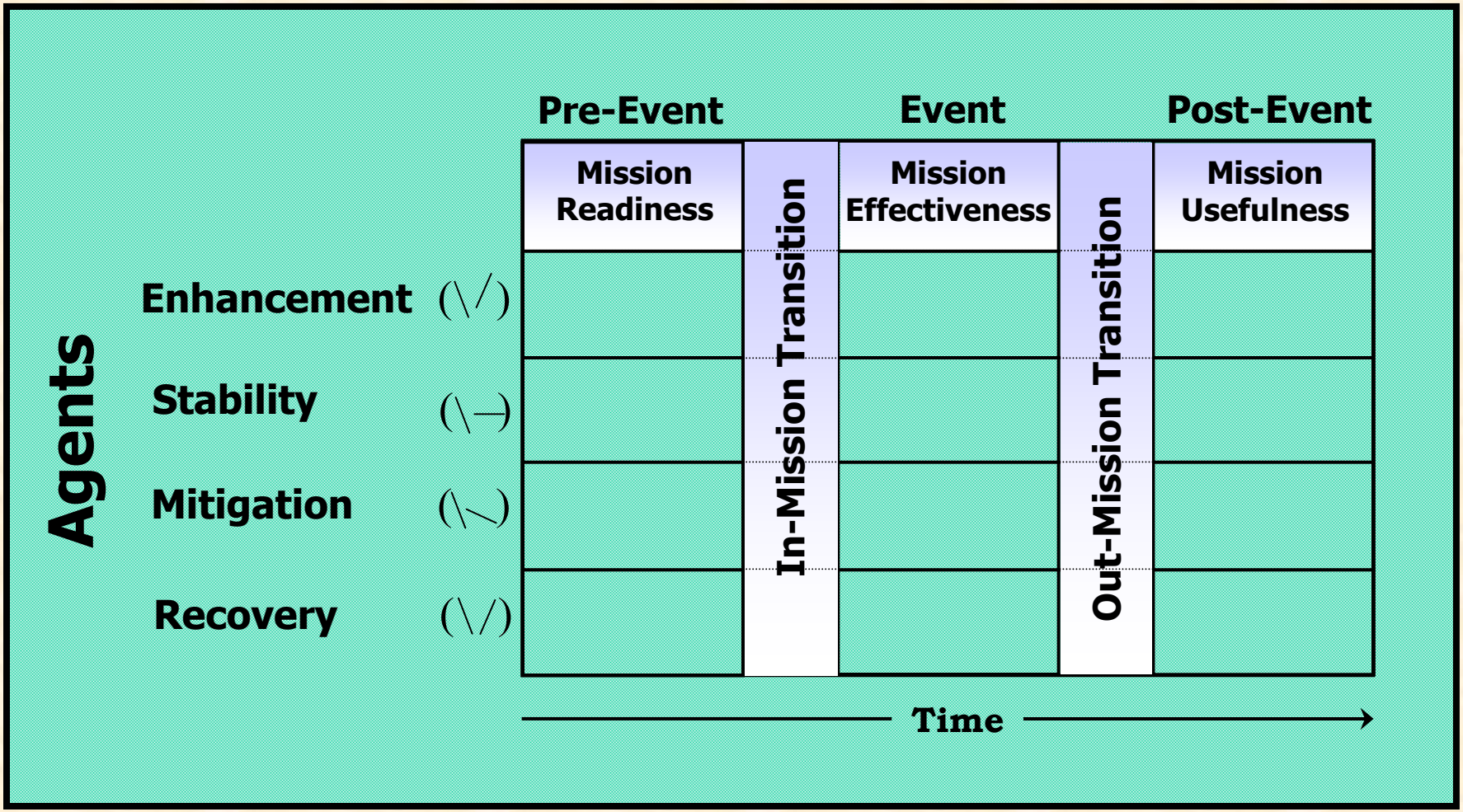
Is Search Memory Disrupted By Noise Stress?

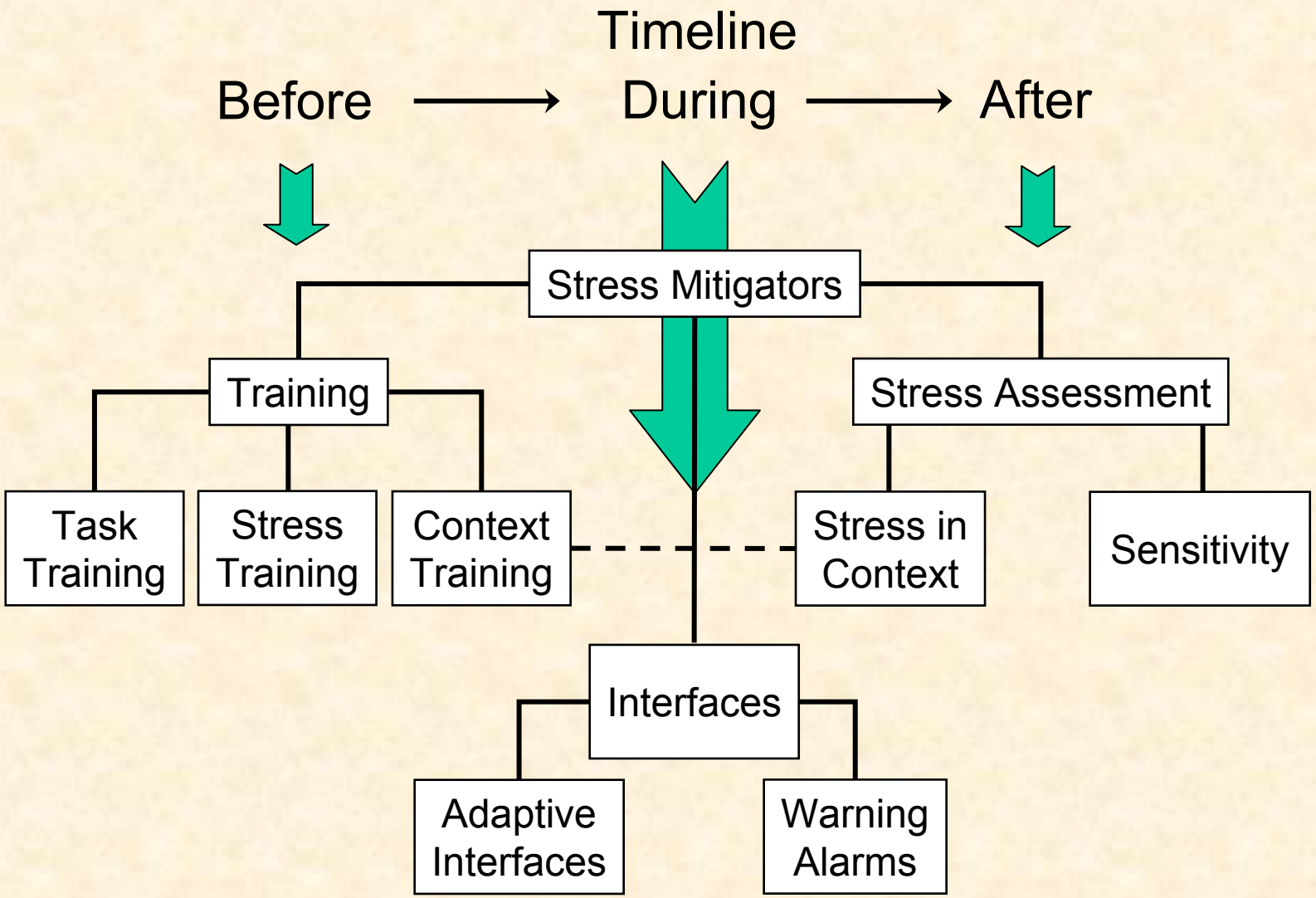




D. D. Diaz. (2002). Visual Search for Two Identical Targets: The Role of Ecological Validity. Annual APA Midyear Symposium. Ft. Belvoir.

Matrix for Performance Change




Timeline
Before
During
After
Stress Mitigators
Training
Stress Assessment
Task Training
Stress Training
Context Training
Stress in Context
Sensitivity
Interfaces
Adaptive Interfaces
Warning Alarms

OPUS Goals

- A fully articulated theory of stress and performance.
- Integration of progress into predictive models.
- Strategies for performance maintenance/enhancement under stress.
- Training approaches for such strategies.
- Interfaces predicated upon such strategic enhancements.

WEBSITE

www.mit.ucf.edu

Operator Performance Under Stress

MURI

MIT²

The Catholic University of America

Institute for Simulation & Training



Peter Hancock, Ph.D.
Department of Psychology
University of Central Florida, IST



Mustafa Mouloua, Ph.D.
Department of Psychology,



Jim Szalma, Ph.D.
Department of Psychology



Clint Bowers, Ph.D.
Department of Psychology, TPL,



Jeanne Weaver, Ph.D.
Department of Psychology



Richard Gilson, Ph.D.
Department of Psychology, CAHFA



Eduardo Salas, Ph.D.
Department of Psychology



Mark E. Koltko-Rivera, Ph.D.
Department of Psychology



Shawn Burke, Ph.D.
Department of Psychology, IST



Raja Parasuraman
Cognitive Science Laboratory
Catholic University



Arthur S. Leon, M.D.
Laboratory of Physiological Hygiene and Exercise Science
University of Minnesota.



Wayne Harris, Ph.D.
Human Factors Research Laboratory
University of Minnesota



Kip Smith, Ph.D.
Department of Psychology Kansas State University



Ava Walker, Ph.D.
Department of Kinesiology
University of Minnesota

Operator Performance Under Stress

MURI

MIT² | The Catholic University of America | Institute for Simulation & Training

UCF Human Factors Psychology Graduate Students


Christina Morris


Adams Greenwood-Ericksen


Krystal Bradley – Radakovich


H.C.Neil Ganey


Jennifer Thropp


Derek D. Diaz


Jessica Helmick


Chris Brill


Lorrenzo Torrez

Undergraduate & Post Bac. Research Assistants


Jennifer Ross


Danielle Adzima


Shawn Stafford


Chrissy Giddens


Craig Russell


Cristina Vega

Coordinator Administrative Services


Barbara Koehler
Coordinator, Administrative Services, UCF

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was Supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Operator Performance Under Stress

MURI

MIT²

The Catholic University Of America

Institute for Simulation & Training

University of Central Florida	Catholic University	University of Minnesota	Kansas State University
MIT²	**IST**		
Time Distortion	Teams Under Stress	Brain States & Stress	Physical & Mental Workload	Decision-Making
Interfaces for Stress	Stress & Team Cognition	Cognitive Workload	Physical Evaluation	Distrust
Visual Search	Low Cost Training	Viking	Viking	Biomedicine
Delta Force	Flashpoint			
Viking				

MURI Affiliations

Dr. Rene de Pontbriand,	ARL, Aberdeen Proving Ground,	Theory and Evaluation
Dr. Laurel Allender,	ARL, Aberdeen Proving Ground,	Modeling and Prediction
Dr. Linda Fatkin	ARL, Aberdeen Proving Ground,	Physical/Cognitive Workload
Dr. Jerry Krueger	The Wexford Group	Soldier Performance
Dr. John Deaton	CHI Systems	Adaptive Interfaces
Dr. Robert Hockey	University of Leeds	Theory and Modeling
Dr. Sherry Tove	ARO	Program Development
Dr. Mike Drillings	ARI	Program Advisement
Dr. Elmar Schmeisser	ARO	Program Direction
Col. Larry Shattuck	West Point	Advisory Council
Col. Bob Tyler	USMC retd	Advisory Council
Col. Bill Blackwood	US Army retd	Advisory Council
Dr. John Ruffner	DCS Inc	Industry Affiliate

Effects of Context-Relevant Stress in PC-Based Game Training

Morris, C.S., Shirkey, E.C., Hancock, P.A., & Mouloua, M.
Department of Psychology
University of Central Florida
Orlando, Florida

ARO DAAD19-01-1-0621

ABSTRACT

This work examined the effects of inserting contextually-relevant stress in a PC-based game for soldier training. Off-the-shelf PC-based games are being applied to multiple training situations because of their affordability, flexibility, and team-training capabilities. One of the major drawbacks to using games as training tools, especially for military applications, is the neglect of surrounding context. In response to this omission, we examined the effects of adding context-relevant stress to infantry game-based training by exposing one group of participants to a graphically intense and stressful war film while the other group viewed an un-stressful analog. Pre-post self-reported stress levels confirmed the efficacy of this manipulation. The stress condition produced significantly higher scores on "mission success" however, no differences were produced in participants' use of trained tactics, or game functions. Supplementing context-relevant stress in game training is consequently recommended to enhance individuals' motivation to succeed.

INTRODUCTION

The simulation-based training movement, with its inherent ability to deliver arousal, offers a promising future for training. A much needed application of this is evidenced by the ability to deliver emergency situations and unexpected danger to the training environment. Simulations, however, specifically developed to provide *entertainment* (the element of gaming), may inadvertently reduce the contextual arousal or stress that learners experience during simulations. Specifically, simulations that are contextually defined as gaming, may threaten training effectiveness in domains that heavily rely on context-relevant stress and intensity as a formal part of the training (e.g. military, emergency responders).
This study attempted to partially rectify the contextual conflicts that are inherent in using entertaining PC-based games for infantry training. Specifically, by supplementing infantry game-based training with an intense and grotesque video depiction of front-line infantry battle, we expected to influence the training and outcome performances of participants.

METHOD

• *Participants:* Participants were 24 student volunteers. An average age of 22 with 15 males and 9 females. No prior military experience and no differences between groups on gaming experience.

• *Procedure & Materials:* A biographical & gaming habit questionnaire was completed, baseline-scores were obtained on the Activation Deactivation Checklist (ADC) (Thayer, 1967) and the Emotional Assessment Scale (EAS) (Carlson, COlins, Stewart, & Porzelius, 1989). A fifteen-minute grotesque segment of the movie Saving Private Ryan was shown to participants in the experimental condition. A fifteen-minute non-grotesque black and white documentary was shown to the control group. Post ADC & EAS scores were obtained then groups were then trained for 20 minutes on soldier tactics, mission objectives, and use of the game. Finally, they were given 20 minutes to try and complete a successful mission. Performance was assessed by a blind observer (game and soldier tactic expert) with a checklist corresponding to the components of training.

RESULTS

TABLE 1.
Post – film reaction differences between groups on indicators of stress/arousal. Groups were not statistically different prior to manipulation.

Stress Indicators	Control Group		Experimental Group		*t*(22)	*P*
	M	*SD*	*M*	*SD*		
ADC						
Clenched	2.0	.51	2.6	.67	-2.05	.05
Intense	2.2	.39	2.7	.65	-2.28	.03
EAC						
Anxiety	2.5	2.03	8.6	8.36	-2.48	.02
Disgust	1.62	1.97	6.94	8.72	-2.05	.05

TABLE 2.
Between-subjects effects of the performance subscales, the DV's, of a significant MANOVA, $F(4,19) = 7.15$, $p = .001$, ($\eta^2 = .60$)

Performance Subscales	Control Group	Experimental Group	*F* (1,22)	*P*
Mission Success	***M* =12.66 *SD*=3.33**	***M* =16.38 *SD* =4.03**	**6.075**	**.022**
Subscale items) 2nd waypoint reached; Avoided taking fire; Use of alternative or creative ways to accomplish mission; Skilled use of time; Followed through with mission and objectives				
Game Adaptation	*M* =24.27 *SD* =3.37	*M* =26.28 *SD* =5.46	.689	.415
Subscale items) Use of provided functions of mouse and keyboard; Switching between weapons; Using full visual (e.g., nightvision, scope) and weapons inventory				
Use of Tactics	*M* =29.50, *SD* =4.47	*M* =28.84 *SD* =7.08	.075	.787
Subscale items) Use of strafing (sidestepping); Cautious approaches into areas; Used scope and maintained distance from targets when engaging; Maintaining cover while engaging				

TABLE 3.
a.) Hierarchical regression model (most parsimonious) predicting "overall performance" from the stress indicators after considering condition assignment.
b.) Stepwise regression model (most parsimonious) predicting performance on the "mission success" subscale when considering both the condition assignment and stress indicators.

	Variables	*B*	β	R^2	Adjusted R^2	*F*	*df*	*p*
a.) Hierarchical Predicting "Overall Performance"	Constant	73.175		.877	.831	19.01	6,16	.000
	Condition	13.626	.36					
	Energetic	40.809	.994					
	Tense	32.302	1.221					
	Clenched	-47.962	-1.447					
	Restful	-11.288	-.427					
	Intense	-13.090	-.400					
b.) Stepwise Predicting "mission success"	(Constant)	6.91		.589	.524	9.07	3,19	.001
	Condition	5.69	.70					
	Energetic	4.14	.47					
	Intense	-2.72	-.39					

DISCUSSION

• The experimental group had significantly higher responses on the self-reported stress indicators than the control group.

• The experimental group had significantly higher scores only on the "mission success" subscale.

  • Interestingly, the "mission success" items (table 2) were not considered as training content, only as goals or mission objectives. Since there were no significant differences between the groups on gaming experience, and initial stress, an interesting conclusion is that context-relevant materials produce motivation to succeed in game training contexts, but does not affect specific skill acquisition.

• Contrary to expectations, "learned tactics", "game adaptation", and "overall performance" (the three subscales combined) did not differ for the two groups.

• To observe individual differences in their stress-reactions on performance we used a hierarchical multiple regression analysis. Individuals' reactions to specific stress indicators, when entered after accounting for the experimental condition, predicted 88% of the variability in overall performance. Performance, then, is best when exposed to the stress condition that results in an individuals' higher motivational energetics (higher energy, higher tenseness, less intensity, and less clenched).

This is the first known study to observe the effects of contextually relevant supplemented material, in the form of added stress, to game-based training. The overall outcome is an increase in motivation on subsequent tasks. It is also suggested that contextually-enhanced material that produces positive energetics, can produce higher outcome training performance.

ACKNOWLEDGEMENT
The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

The Effect of Noise Stress During Search for Two Identical Targets in a Realistic Visual Scene

Derek D. Diaz, Peter A. Hancock, Valerie K. Sims, Christopher Brill
University of Central Florida

Research Question

Is search memory disrupted by noise stress?

- During a complicated visual search, eye-movements are guided by memory from previous locations and objects, or "search memory".
- This automatic process facilitates efficient search by guiding attention to novel stimuli.
- The presence of noise stress can affect search performance.

Introduction

Visual Search in the Real World

Visual search is part of our daily lives!

- Examining an x-ray image for an anomaly
- Searching for a character on a keyboard
- Searching for a robin in a tree
- Security surveillance
- Scanning for a street sign

For concerns of validity, previous research has focused on stimuli consisting of singletons, letters, numbers, and esoteric symbols (see examples provided at right). This study is unique in that it examines visual search of a realistic scene. This was made possible by using 3D modeling technology to create realistic human figures which served as search stimuli.

Search Memory Aids Visual Search

- Search memory is a low-level, automatic process
- Search memory guides visual attention to new objects and facilitates efficient search
- Search memory has been found to be more robust when the objects being searched reflect real life objects (Diaz, Hancock, & Sims, 2002)

The present study manipulated the ability to use search memory by continuously repositioning targets and distractors into random locations

Noise Stress Affects Search Performance

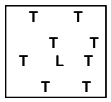
Previous research has found that noise stress can have different effects on visual search

- Increase search time and increase errors (e.g., see Hockey & Hamilton, 1986)
- Decrease search time and decrease errors (e.g., see Orasanu & Becker, 1996)

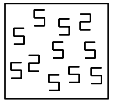
The present study examines whether the effects of noise on visual search occurs, in part, by disrupting search memory. Visual search is examined in the context of a realistic scene (i.e., people running). Our previous work has found that search memory is more robust under such conditions.

Stimuli From Previous Visual Search Experiments

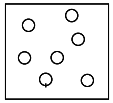
T's & L's



Gibson, Li, Brown, Skow, and Cooke, 2000



Klein, 1988



Wolfe, 1993



Stimuli Used in the Present Study

Target



Distractors



Independent Variables

2 x 2 x 2 Mixed Factorial Design

Target-type (within)
Single vs. Double

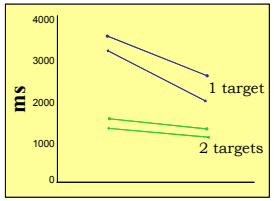
Search-type (within)
Dynamic (random)
Static



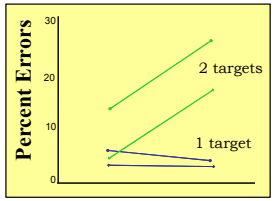
85 db White Noise (between): Present vs. Absent

Results

Response Time



Response Accuracy



No Yes No Yes

Presence of Noise Stressor

Dynamic-1 target
Dynamic-2 targets
Static-1 target
Static-2 targets

Method

The search task was based on the procedure used by Gibson, Li, Brown, Skow, and Cooke (2000) and required observers to determine via button push if one or two identical targets were present among an array of distracters.

To examine performance when search memory is both functional and impaired, the search objects followed random movement patterns for half of the trials.

Participants

40 undergraduates (20 females) were randomly assigned into the stress conditions (noise present vs. noise absent).

Design

- Independent Variables:
  - Target-type (within): single vs. double targets
  - Search-type (within): dynamic (random vs. static
  - Noise stress (between): present vs. absent
- Dependant Variables:
  - Response Time
  - Accuracy

Procedure

- 400 trials projected onto 6' x 4' screen
- 15 total search objects
  - Either 1 target & 14 distracters
  - Or 2 targets & 13 distracters
- 4-second white noise at 85 dB was presented on random trials in the stress condition
- Response via one of two push-buttons (for 1 or 2 targets)
- Participants had to decide if there were 1 or 2 targets present in each trial

Discussion

The effect of noise stress depended on the number of targets present

- When 1 target was present:
  - Response time decreased
  - Accuracy was unaffected
- When 2 targets were present:
  - Response time was unaffected
  - Accuracy decreased
- A general implication from these findings is that, under environmental conditions with loud and unpredictable noise, search for multiple targets is more challenging than search for a single target.
- As indicated by the increase in response errors when 2 targets were present, results suggest that the presence of a noise stressor has a negative effect on visual searches that require search memory.

References

Gibson, B. S., Li, L., Skow, E., Brown, K., & Cooke, L. (2000). Searching for one versus two identical targets: When visual search has a memory. *Psychological Science, 11*, 324-327.

Horowitz, T.S. & Wolfe, J. M. (1998). *Visual Search has no memory. Nature, 357*, 575-577.

Klein, R. M. (1988). Inhibitory tagging systems facilitates visual search. *Nature, 334*, 430 – 431.

Klein R. & MacInnes, W.J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, 10, 346-352.

Kristjansson, A. (2000). In search of Remembrance: Evidence for memory in visual search. *Psychological Science, 11*, 328-332.

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science, 12*, 287-292.

A Multi-sensory Device and Protocol for Online Workload Assessment

Brill, J. C.*, Mouloua, M.*, Hancock, P. A.*, Kennedy, R. S., Gilson, R. D.*, & Smith, M. G.****

* University of Central Florida, ** RSK Assessments, Inc.

Introduction

- No existing measures of workload assess cognitive resource pool capacities across sensory modalities or across tasks.
- Primary measures of workload are task specific.
- Other workload measures can only be obtained retrospectively (e.g., NASA-TLX).
- To circumvent these limitations, the Multi-Sensory Workload Assessment Protocol was developed (M-SWAP).

Design Goals

- Five primary objectives were identified when developing the M-SWAP:
  1. Portability
  2. Independence from language requirements and cultural biases
  3. No special abilities required (e.g.,mathematical computation or reading)
  4. Easy to learn; no practice effects
  5. Relatively inexpensive to build and maintain

Advantages of M-SWAP

- Can be used to determine differential demands of a task on cognitive resources
- Can assess resources across sensory modalities
- Can perform online workload assessment, meaning it can give you real-time (or near real-time) measures of workload
- No task specificity

M-SWAP: How it Works?

- Same methodology, regardless of modality
- Participants presented with series of three randomly presented targets:
  - Visual Task: Three flashing lights (left, center, right)
  - Auditory Task: Three auditory pitches (low, middle, high)
  - Tactile Task: Three loci of vibration (left, center, right)
- Participants count frequencies, hit button every four count
- Counting errors serve as measure of workload (e.g., high number of errors = high workload)
- Difficulty of M-SWAP is scalable



Demonstration of desktop driving simulator



Portable LCD Display for the Visual Counting Component of M-SWAP



In-car navigation device and hands-free mobile phone used for on-road testing

Apparatus

- M-SWAP hardware consists of the following:
  - Wearable vibrotactile display
  - 2.5 Watt Amplifier
  - PC-controlled Relay Board
  - Function Generator
  - Personal Computer with a Portable LCD Monitor
  - 3-button Response Box
  - Proprietary Software Package



Imtech® Model GFG-8210 Function Generator



Engineering Acoustics Inc., Model C2 Tactor
- shown with a quarter for size comparison -



Tactor Controller Box with Velotrack® Tactor Belt and C2 Tactors

Want to learn more?

For more information on the M-SWAP, feel free to contact us via e-mail:

Chris Brill
jbrill@pegasus.cc.ucf.edu

Mustapha Mouloua
mouloua@pegasus.cc.ucf.edu

This research was facilitated by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ARO under grant DAAD19-01-1-0621. The conception and development of this research was also funded by the AAA Foundation.

M-SWAP Evaluation: Preliminary Results

- Protocol was tested using a desktop driving simulator and an on-road vehicle
- Driving in light traffic under three conditions:
  - **Dual Task**: driving + visual counting *
  - **Tertiary Task A**: driving + operating in-car navigation system + visual counting
  - **Tertiary Task B**: driving + talking on cell phone + visual counting

* Only the single target, low difficulty counting task was used.

Pilot Study Results

Condition	*Mean Errors per 4 minutes on task*
Driving + Visual Counting	**1.0**
Driving + Visual Counting + Navigation System	**2.5**
Driving + Visual Counting + Cell Phone	**4.5**

Discussion

- Results indicate the M-SWAP is sensitive to differential levels of workload.
- Results also suggest that use of while driving telematic devices is attentionally demanding and reduces resource capacity, but to varying degrees.
- Current Work:
  - To establish the convergent and divergent validity of the M-SWAP
  - To test the M-SWAP in a variety of situations under varying workload levels
- Also using M-SWAP for:
  - Evaluating driver distraction
  - Evaluating stress and workload in combat operations
  - Cockpit design

A MULTI-SENSORY DEVICE AND PROTOCOL FOR ONLINE WORKLOAD ASSESSMENT

Brill, J. C.*, Mouloua, M.*, Hancock, P. A.*, Kennedy, R. S.**, Gilson, R. D.*, & Smith, M. G.**

* University of Central Florida, ** RSK Assessments, Inc.

Although a variety of workload measures exist, none of the existing measures are capable of assessing the capacities of cognitive resource pools across multiple sensory modalities and across various tasks. The goal of the present effort was to develop a device to help assess the workload associated with complex tasks involving more than one sensory system. Five design criteria were established to guide the development of the workload assessment device and protocol. First, portability was deemed essential to maximize the utility of the device across research scenarios, regardless of whether it is used in the field or in the laboratory. Second, the protocol should be independent of language requirements and cultural bias. Third, use of the device should not require special abilities such as mathematical computation or reading. Fourth, the protocol should be easy to learn, and there should be an absence of practice effects. Fifth, the device should be inexpensive to build and maintain to increase its accessibility. It is believed that the workload assessment device and protocol envisioned below meets all five of these criteria.

The workload assessment device consists of a wearable display as an input to the least used sensory modality: touch. The system is comprised of three vibrotactile actuators (called tactors), a relay-based tactor controller board, a three-button response box, and a PC using proprietary software developed by the research team. The system is used for presenting a multi-sensory counting task for assessing workload while operators perform complex tasks (such as driving while using telematic devices including cell phones, in-car navigation systems, or even tuning a radio).

The protocol for using the workload assessment device is generally the same regardless of which sensory modality (i.e., visual, auditory, or tactile) is targeted, but the following description depicts the visual component of the protocol. Participants are presented with a series of lights flashing randomly at three locations (left, center, or right) on a computer screen. They will count the frequencies of lights (each four count), and respond by pressing one of three response buttons. The number of counting errors serves as a measure of workload, wherein low workload tasks should be associated with few counting errors and high workload tasks should be associated with greater frequencies of counting errors. The difficulty of the counting task is scalable depending upon the number of stimulus sites subjects are instructed to monitor (one site = low difficulty, two sites = moderate difficulty, etc.). The counting task for the other two sensory modalities is similar wherein subjects are counting from among either three loci of vibrotactile stimulation (left side of torso, midline, or right side of torso) or three auditory pitches (low, middle, and high). The counting task is an adaptation of the one used by Kennedy and Bittner (1980).

Aside from meeting the five aforementioned design criteria, there are other significant advantages of the multi-sensory workload assessment protocol (hereafter referred to as "M-SWAP") over other established workload measures. For example, the NASA Task Load Index (NASA-TLX) is a commonly used subjective self-report workload measure in which operators rate items on several task dimensions such as mental demands, temporal demands, effort, and frustration (Hart & Staveland, 1988). It is limited in that it cannot be administered while someone is performing a task, as can the M-SWAP. In addition, the NASA-TLX treats cognitive resources as a unitary pool (e.g., Broadbent, 1958), which restricts its utility for situations in which one wishes to determine the differential demands of a task on cognitive resources, especially if one's research is rooted in theories purporting

the existence of multiple resource pools of attention (e.g., Wickens, 1984). In fact, assessing resources across and within modalities is a distinct advantage of the M-SWAP system. With the M-SWAP, workload can be assessed on-line (meaning while the operator performs the task) and in real-time, without the limitations of task-specificity associated with primary task measures (e.g., Linton's (1975) statistical workload assessment model (SWAM); North & Riley's (1988) Workload Index (W/INDEX). This real-time (or nearly real-time) assessment is essential for the design of adaptive technologies in which elements of a complex task might be automated through feedback if excessive workload levels are detected. Finally, the secondary task paradigm used by the M-SWAP can be used to determine the availability of resources across sensory modalities. These data can also be used for system design to maximize the likelihood of signal detection while performing complex tasks. For example, one might use the M-SWAP for confirmation of the availability of resources within a sensory modality that is theoretically untaxed by a task to aid in making design decisions (e.g., to what sensory modality should an alarm system target).

Testing to establish the validity of the M-SWAP is continuing, but initial evaluation of the device and protocol in both a driving simulator and in an on-road vehicle suggests the M-SWAP is sensitive to different levels of workload. For example, while operating an automobile in light traffic participants committed an average of one counting error (visual) every four minutes. When given the additional task of either operating a navigation system or talking on a cell phone, mean counting errors increased 2.5 and 4.5 per four-minute block, respectively. These results suggest that the M-SWAP is sensitive to differential levels of workload.

Selected References

Broadbent, D. (1958). *Perception and Communication*. New York: Pergamon.

Hart, S., & Staveland, L. (1988). Development of the NASA-TLX: Results of empirical and theoretical research. In P. Hancock and N. Meshkati (eds.), *Human Mental Workload*. Amsterdam: North-Holland.

Kennedy, R. S., & Bittner, A. C. (1980). Development of performance evaluation tests for environmental research (PETER): Complex counting test. *Aviation, Space, & Environmental Medicine*, 51(2), 142-144.

North, R., & Riley, V. (1988). W/INDEX: A predictive model of operator workload. In G. MacMillian (ed.), *Human Performance Models*. Orlando, FL: NATO AGARD Symposium.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of Attention* (pp. 63-101). Hillsdale, NJ: Erlbaum.

For more information about the M-SWAP, feel free to contact us:

Chris Brill
jbrill@pegasus.cc.ucf.edu

Mustapha Mouloua
mouloua@pegasus.cc.ucf.edu



Multi-Disciplinary University Research Initiative

MIT²

The Catholic University Of America

INSTITUTE for SIMULATION & TRAINING



The Influence of Stress and Individual Differences on the Perception of Duration

J.L.Weaver & P.A. Hancock

ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.



"In general, a time filled with varied and interesting experiences seems short in passing, but long as we look back. On the other hand, a tract of time empty of experiences seems long in passing, but in retrospect short."

William James, 1890

Background

- Paradoxical phenomenon
- Study of *perceived time*, or *protensity*, has been conducted from the perspective of a number of different disciplines (e.g., philosophical, biological, psychological) (Fraser, 1987).
  - Endogenous (i.e., "the internal clock", cf. Hoagland, 1933) vs. Exogenous Factors studied in relation to time perception

IMPETUS FOR CURRENT RESEARCH

- Relatively little is understood about time perception under *stress*.
- Both inter- and intra-individual differences exist with respect to time perception.
- Limited study of the role of individual differences (e.g., personality, age, culture) on perceived time (cf., Doob, 1971; Friedman, 1990).
  - Found that there are individual differences in the way that people *think* about time and its meaning.
- Better understanding of time distortions has very real and practical applications.

CHALLENGES

- Methodological Issues
  - Research in this area is limited by the elusive nature of the phenomenon (does not appear on demand).
  - Turning subjective experience into externally observable actions/metrics also presents a significant challenge (see Hancock, Weaver, & Parasuraman, in press).
  - Manipulating stress in the laboratory

Attentional Resource Theory

Time distortions occur due to the depletion of attentional resources available as a function of stress. Specifically, stress limits resources available to attend to environmental *and* time-based cues (see Hancock & Weaver, in press).

Current Study

Purpose

- To test whether different temporal instruction sets would alter the perception of time-passage.
- To test whether these responses would be moderated by individual difference characteristics.
  - Need for Cognition (Cacioppo & Petty, 1982).
  - Self-Control (Rosenbaum, 1980).

Hypotheses

- Perceived time passage would be longer in the high time-pressure condition.
- Time perception would be more accurate in the low time-pressure condition.
- Time perception would be more accurate in high self-control participants.
- Time perception would be more accurate in low need for cognition participants.

Hypotheses

- Time estimations would be most accurate for low need for cognition, high self-control participants exposed to low time-pressure.
- Time estimations would be least accurate for high need for cognition, low self-control participants exposed to high time-pressure.

Method

Participants

- Male and female undergraduate psychology students enrolled at UCF. Ss were offered extra credit in exchange for their participation.
- Participants were randomly assigned to either the high or low time-pressure condition.

Materials

- Used a driving simulation, displayed on a 17" computer monitor fitted with a steering wheel and pedal input device system.

Method

- Participants completed a demographic questionnaire, the ***Need for Cognition Scale,*** and the ***Self Control Schedule*** upon arrival.
- Ss were given task practice.
- They were told they would be driving in a simulated driving environment, to follow all normal traffic laws, and to practice accelerating, braking and turning.
- During the test period, participants were asked to read one of two scenario sheets (low vs. high time-pressure) and asked to drive the course highlighted by the experimenter and that also appeared at the bottom right hand corner of the dashboard display.



Method

- Ss were all stopped at the same point in the scenario and presented with a response sheet asking them to indicate how much time they believed had elapsed during the final red light (time perception) and how much time pressure they felt during the scenario (manipulation check).

Results

- Method of verbal estimation
  - all estimates given in round numbers
- Manipulation Check
  - Participants perceived more time-pressure in the "high time-pressure" condition than in the low "time-pressure" condition, $t(119)= -8.111$, $p<.005$.
- Time estimate variability
  - Range = 2 to 120 sec.
  - Standard deviation = 14.87

Results

- Hypothesis Tests
  - To test the hypothesis that time estimations would differ as a function of time-pressure condition, a one-way between subjects ANOVA was conducted with time estimation as the dependent variable.
    - There was no significant difference between the time-pressure groups ($p > .05$).

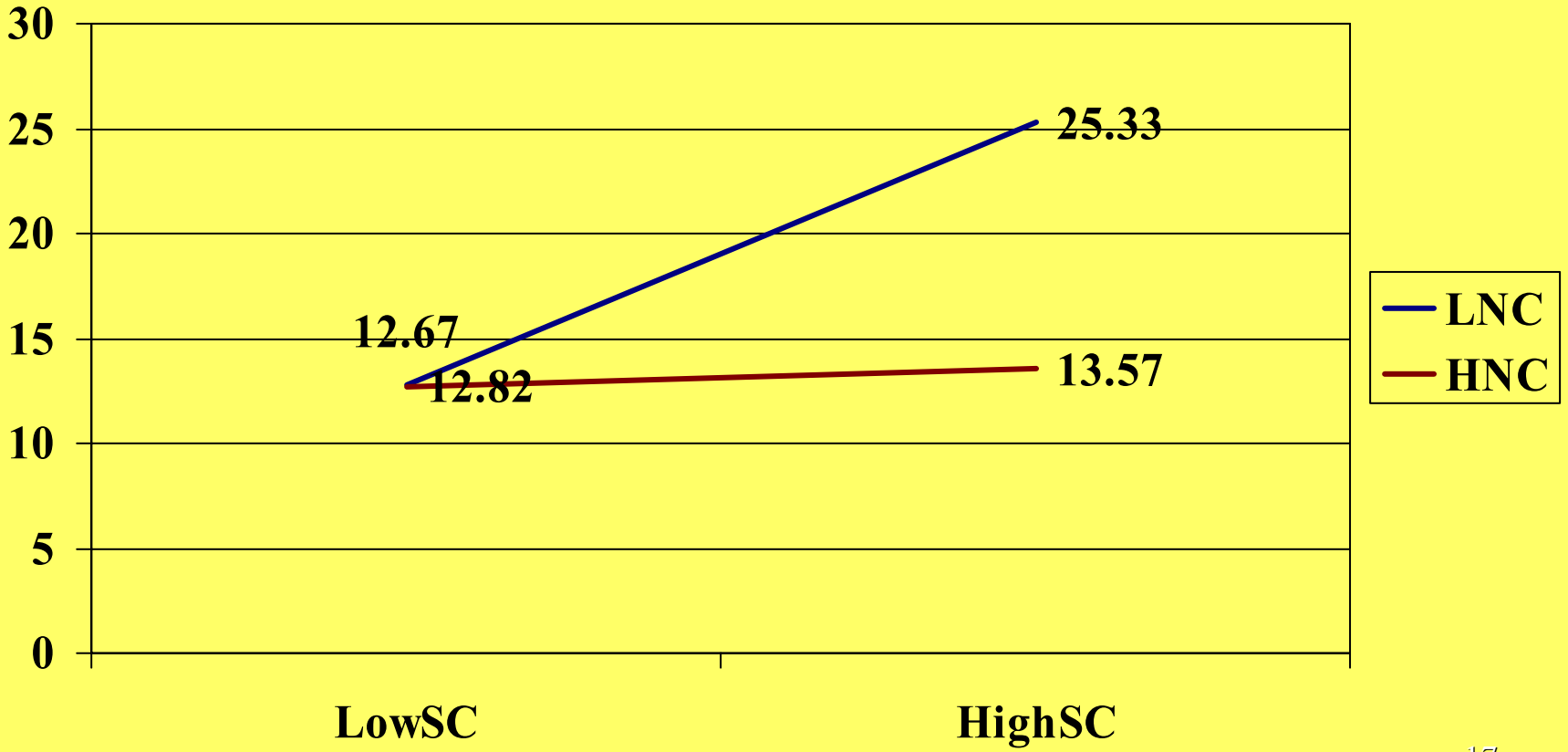
Results

- Hypothesis Tests
  - Conducted a 2 x 2 x 2 between subjects ANOVA with time deviation as the dependent variable.
    - Significant main effect of self-control, with high self-control persons more accurate (M = 19.45) than low self-control (M = 12.75) (median split) $F(1, 90)= 4.531$, $p=.036$.
    - Marginal main effect of need for cognition, with low need for cognition persons more accurate (M = 19.003) than high (M = 13.2) (median split) $F(1, 90)= 3.402$, p=.064.

Results

- Hypothesis Tests (Cont.)
  - Conducted a 2 x 2 x 2 between subjects ANOVA with time deviation as the dependent variable.
    - Marginal interaction of need for cognition and self-control persons with high self-control, low need for cognition persons being most accurate, followed by high self-control/high need for cognition, high need for cognition/low self-control, and low need for cognition/low self-control $F(1, 90)= 3.58$, $p=.062$.

Self-Control & Need for Cognition Interaction



Results

- Manipulation Check
  - Participants perceived more time-pressure in the "high time-pressure" condition than in the low "time-pressure" condition, $t(147) = -9.582$, $p < .001$.
- Hypothesis Tests
  - Conducted a 2 x 2 x 2 between subjects ANOVA with time deviation as the dependent variable.
    - Significant main effect of self-control, with high self-control persons more accurate ($M = 18.60$) than low self-control ($M = 13.14$) (median split) $F(1, 135) = 4.346$, $p = .039$.
    - No other significant effects ($p > .05$)

Discussion

- Overall underestimation of time-passage
- Great deal of variability in estimates
- Perceived time-passage varies as a function of individual difference variable (self-control)
  - Nature of variables and attentional resources

Discussion

- Question the stress manipulation (salience)
  - Time-pressure as a component of the task vs. given as instructions
- Still need to determine role of stress in affecting perceived time passage
  - Multi-tasking in multi-modes
  - Field investigations (e.g., EMS)

Implications

- Training
  - Can people be trained to be more accurate in time perception?
- Selection
  - Test whether there are benefits associated with the overestimation of time-passage.
    - Accomplish more in “less time”
- Technology
  - Technical aids to increase time-cue salience

Multi-Disciplinary University Research Initiative

The Influence of Stress and Individual Differences on the Perception of Time

J.L.Weaver, V.K. Sims, H.C.N. Ganey,
P.A. Hancock, K.A. Bradley, & M. Mouloua,



ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Mutidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Background

- Study of *perceived time*, or *protensity*, has been conducted from the perspective of a number of different disciplines (e.g., philosophical, biological, psychological) (Fraser, 1987).
- Relatively little is understood about time perception under stress.
- Limited study of the role of individual differences (e.g., personality, age, culture) on perceived time (cf., Doob, 1971; Friedman, 1990).
  - This area of research has found that there are individual differences in the way that people *think* about time and its meaning.

Background

- It has been argued that stress depletes resources thus limiting available attention to attend to environmental *and* time-based cues (see Hancock & Weaver, under review).
- Methodological Issues
  - Research in this area is limited by the elusive nature of the phenomenon (does not appear on demand).
  - Turning subjective experience into externally observable actions also presents a significant challenge (see Hancock, Weaver, & Parasuraman, in press).

Current Study

Purpose

- To test whether different temporal instruction sets would alter the perception of time-passage.
- To test whether these responses would be moderated by individual difference characteristics.
  - Need for Cognition (Cacioppo & Petty, 1982).
  - Self-Control (Rosenbaum, 1980).

Hypotheses

- Perceived time passage would be longer in the high time-pressure condition.
- Time perception would be more accurate in the low time-pressure condition.
- Time perception would be more accurate in high self-control participants.
- Time perception would be more accurate in low need for cognition participants.

Hypotheses

- Time estimations would be most accurate for low need for cognition, high self-control participants exposed to low time-pressure.
- Time estimations would be least accurate for high need for cognition, low self-control participants exposed to high time-pressure.

Method

Participants

- Male and female undergraduate psychology students enrolled at UCF. Ss were offered extra credit in exchange for their participation.
- Participants were randomly assigned to either the high or low time-pressure condition.

Materials

- Used a driving simulation, displayed on a 17" computer monitor fitted with a steering wheel and pedal input device system.

Method

- Participants completed a demographic questionnaire, the ***Need for Cognition Scale,*** and the ***Self Control Schedule*** upon arrival.
- Ss were given task practice.
- They were told they would be driving in a simulated driving environment, to follow all normal traffic laws, and to practice accelerating, braking and turning.
- During the test period, participants were asked to read one of two scenario sheets (low vs. high time-pressure) and asked to drive the course highlighted by the experimenter and that also appeared at the bottom right hand corner of the dashboard display.



Method

- Ss were all stopped at the same point in the scenario and presented with a response sheet asking them to indicate how much time they believed had elapsed during the final red light (time perception) and how stressful they felt the scenario was (manipulation check).

Results

- Manipulation Check
  - Participants perceived more time-pressure in the "high time-pressure" condition than in the low "time-pressure" condition, $t(119)= -8.111$, $p<.005$.
- Time estimate variability
  - Range = 2 to 120 sec.
  - Standard deviation = 14.87
- Method of verbal estimation gave all estimates in round numbers

Results

- Hypothesis Tests
  - To test the hypothesis that time estimations would differ as a function of time-pressure condition, a one-way between subjects ANOVA was conducted with time estimation as the dependent variable.
    - There was no significant difference between the time-pressure groups ($p > .05$).

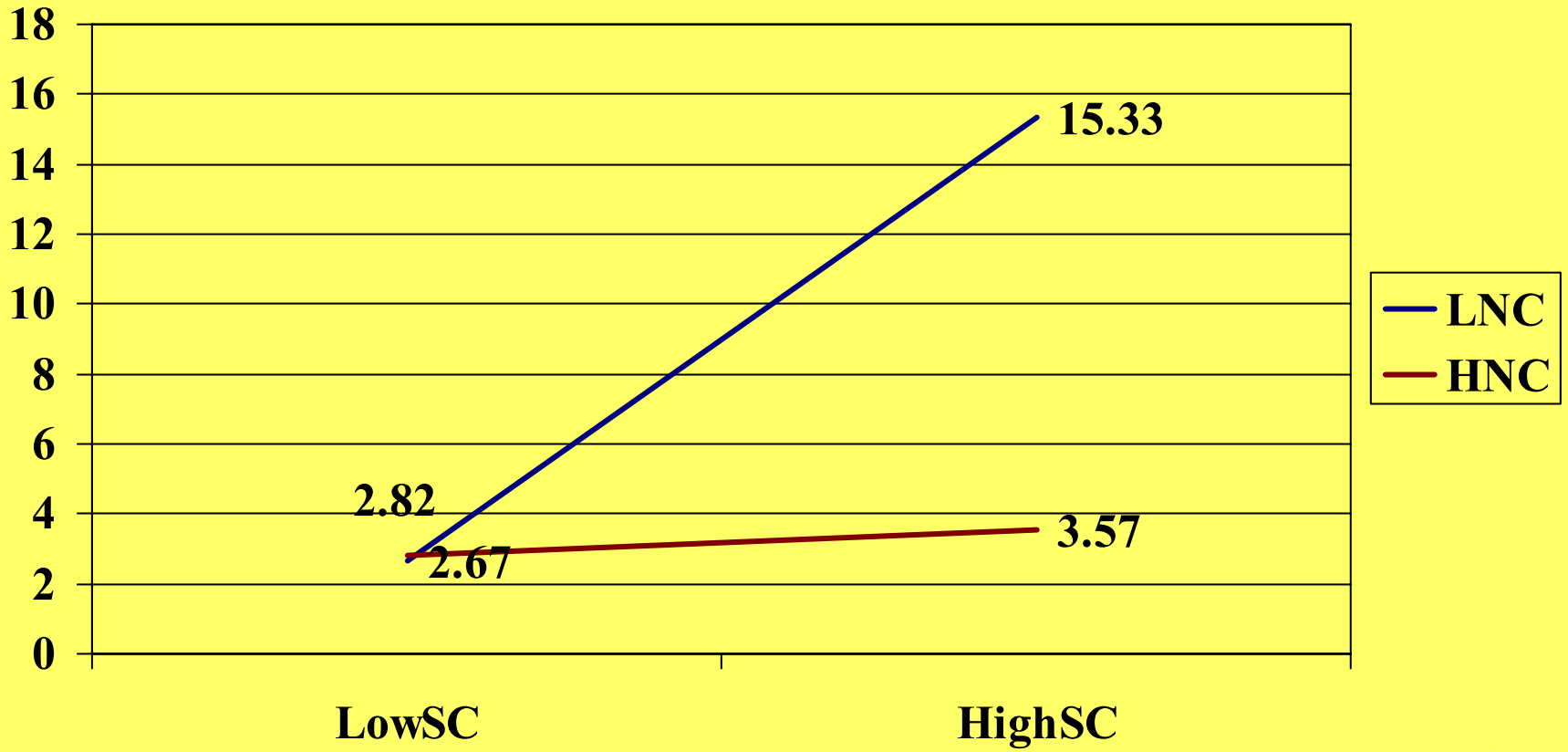
Results

- Hypothesis Tests
  - Conducted a 2 x 2 x 2 between subjects ANOVA with time deviation as the dependent variable.
    - Significant main effect of self-control, with high self-control persons more accurate ($M = 9.45$) than low self-control ($M = 2.75$) (median split) $F(1, 90) = 4.531$, $p = .036$.
    - Marginal main effect of need for cognition, with low need for cognition persons more accurate ($M = 9.003$) than high ($M = 3.2$) (median split) $F(1, 90) = 3.402$, $p = .064$.

Results

- Hypothesis Tests (Cont.)
  - Conducted a 2 x 2 x 2 between subjects ANOVA with time deviation as the dependent variable.
    - Marginal interaction of need for cognition and self-control persons with high self-control, low need for cognition persons being most accurate, followed by high self-control/high need for cognition, high need for cognition/low self-control, and low need for cognition/low self-control $F(1, 90)= 3.58$, $p=.062$.

Self-Control & Need for Cognition Interaction



Discussion

- Overall underestimation of time-passage
- Great deal of variability in estimates
- Perceived time-passage varies as a function of individual differences (self-control and need for cognition)
  - Nature of variables and attentional resources
- Question the stress manipulation
  - Still need to determine role of stress in affecting perceived time passage
  - Time-pressure as a component of the task vs. given in instructions

Implications

- Training
  - Can people be trained to be more accurate in time perception?
- Selection
  - Test whether there are benefits associated with the overestimation of time-passage.
    - Accomplish more in “less time”

Visual Search For Two Identical Targets: The Effect of Noise Stress

Derek Diaz; Faculty sponsors: P.A. Hancock, V.K. Sims

University of Central Florida

Research Question

Is "search memory" disrupted by noise stress?

Background

Visual search of some kind is an important component of most tactical missions

- E.g., targeting, scouting, and reconnaissance
- Threats and targets move or hide

Search Memory

- Memory for items that have already been searched
- Facilitates efficient search
- Guides visual attention to new objects

Klein, R. M. (1988). Inhibitory tagging systems facilitates visual search. *Nature*, 334 (4), 430 – 431

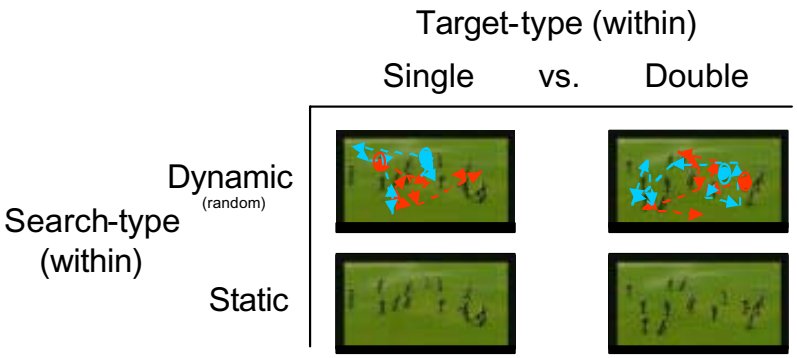
Purpose

The present research explores how memory in visual search is affected by stress and the ecological validity of the target objects. Ecological validity here refers to the degree to which stimuli in the observer's environment obey the perceptual "rules" of the world.

Previously, we have shown that the difficulty of certain types of visual search (as seen in studies using randomly moving geometric forms) can be alleviated when the items to be searched are human forms moving along random trajectories (Diaz, Sims, & Hancock, 2002). This finding has implications for the design of augmented environments which use geometric forms to represent important targets. Such attempts to alleviate workload may actually induce greater effort and result in lower response accuracy.

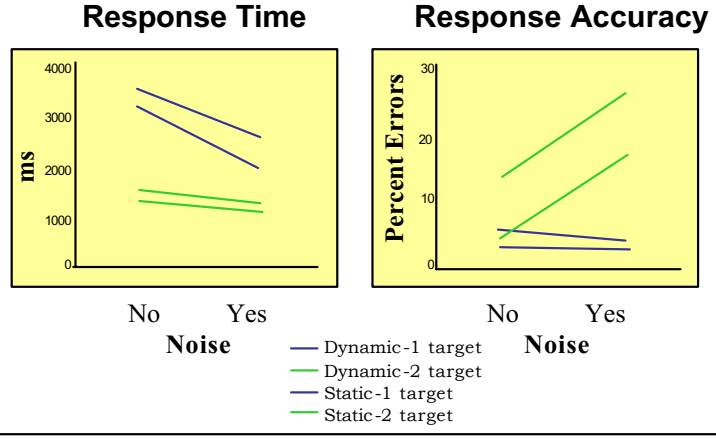
The present experiment extends this research by examining the effect of an environmental stressor on visual search performance. Specifically, we compared performance on an ecologically-valid search task with and without the presence of unpredictable bursts of white noise. Such conditions begin to provide an analog for visual search in oppressive and stressful environments.

Independent Variables



Noise (between): Present vs. Absent

Results — Speed/Accuracy Tradeoff



Hypotheses & Method

We anticipated that random bursts of white noise will decrease reaction time compared to the same task without the noise stress or. We expect these results will occur due to increased attentional focus on relevant cues.

The search task was based on the procedure used by Gibson, Li, Brown, Skow, and Cooke (2000) and required observers to determine via button push if one or two identical targets were present among an array of distracters.

To examine performance when search memory is both functional and impaired, the search objects followed random movement patterns for half of the trials.

Participants

20 undergraduates (10 females) were randomly assigned into the stress conditions (noise present vs. noise absent)

Procedure

- 400 trials projected onto 6' x 4' screen
- 15 total search objects
  - Either 1 target & 14 distracters
  - Or 2 targets & 13 distracters
- Response via one of two push-buttons (for 1 or 2 targets)
- 4-second white noise at 85 dB was presented on random trials in the stress condition

Discussion

Search rate proceeded as expected for a serial terminating search, in that, observers were slower to confirm the presence of two targets than to confirm the presence of only one. In light of previous research finding that random movement of iconic symbols interferes with the use of search memory (Gibson et al, 2001), these findings suggest that the ecological nature of the stimuli used here may mediate the efficacy of search memory.

Overall, noise-stress decreased reaction time as was expected. When noise was present, response time for single-targets increased a greater degree than did response time for double-targets. Moreover, a speed/accuracy tradeoff was not observed for the single-targets while the error rate for the double-targets tripled under noise stress. These findings indicate that observers were conservative in responding to confirm the presence of only single targets and are able to perform at the same level when under stress. These findings further suggest that the speed/accuracy tradeoff during search for double-targets has a detrimental affect on accuracy while producing only a slight decrease in response time.

A general implication of these findings is that, under environmental conditions with loud and unpredictable noise, search for multiple targets is more challenging than search for a single target. Moreover, ecologically-valid movement may produce additional cues that facilitate the detection of two identical targets. Detection of two identical static targets was associated with a substantially higher error rate. These effects were observed both with and without the presence of noise-stress.

Klein, R. M. (1988). Inhibitory tagging systems facilitates visual search. *Nature*,334 (4),430 –431

Klein, R.M. &Macinnes , W.J. *Inhibition of Return is a Foraging Facilitator in Visual Search*.Psychological Science, July 1999

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley , J. S. (2001). Visual search has memory. *Psychological Science*,12, 287 - 292

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science*,12, 219 - 224

Bowers C. A., Weaver, J. L. & Morgan, B. B. (1996). Moderating the performance effects of stressors. In J. A. Driskell & E. Salas (Eds.), *Stress and Human Performance* (pp. 49 -86).Mahwah, NJ: Lawrence Erlbaum Associates.

Gibson, B. S., Li, L., Skow, E., Brown, K., & Cooke, L. (2000). Searching for one versus two identical targets: When visual search has a memory. *Psychological Science*, 11, 181 - 186

Horowitz, T.S. & Wolfe, J. M. (1998). *Visual Search has no memory*. *Nature* , 357, 575-577

The Effect of Time Stress on Mission Performance in a Virtual Combat Mission: Preliminary Results

Adams Greenwood-Ericksen and H.C.N. Ganey.
Faculty sponsors: J.L. Weaver, P.A. Hancock, and M. Mouloua.
University of Central Florida

ABSTRACT

In today's technical world, many individuals often encounter situations in which they must make decisions under time pressure. This is particularly true in the context of combat, where soldiers encounter highly stressful conditions. The purpose of this ongoing study is to examine the effects of time stress on the performance of participants engaged in a simulated mission environment. We proposed that certain indices of individual performance would be degraded in the presence of time-pressure. To date, ten undergraduate subjects have completed hour-long training courses in the use of a soldier simulator based on a commercially-available video game, and then completed a simulated POW rescue mission under one of two time-pressure conditions. In the control condition participants experienced no time stress, while participants in the experimental condition performed the mission under time pressure. Dependent variables fell into two categories, subjective and objective. Subjective DVs consisted of self assessments by the participant of the time pressure they experienced, their performance on the task, and the difficulty of the task. Objective DVs were participant shooting accuracy, number of enemy kills, number of "hits" taken, number of hostages rescued, and time taken to complete mission objectives. At this time, only the subjective data has been analyzed, and these results are presented below in the results section. Our findings and their implications for infantry combat performance under conditions of time stress and information load are discussed.

INTRODUCTION

Combat is most likely the most stressful task performed by human beings, from both a physical and a physiological perspective. Moreover, the consequences of inadequate performance in this role can be catastrophic not only for the individual soldier, but for the unit, army, or even the nation as a whole. Because of its inherently stressful nature and the vital importance of successful outcomes in combat, it is vital that an experimental examination of all factors affecting basic soldier performance be performed in order to gain a more complete picture of soldier performance under stress. To this end, a between subjects study was conducted to investigate participant performance on a low-fidelity soldier simulator (a commercially-available video game called Delta Force: Landwarrior by Novalogic) under varying levels of time stress. Time stress has been previously established as a valid source of stress (Hancock & Desmond, 2001; Doob 1971)

The purpose of the current study is to examine the effects of perceived time pressure on the performance of soldier operators engaged in a simulated mission environment. We predicted that certain indices of individual performance would be degraded under conditions of significant time-pressure and high levels/inappropriate types of information load (the IV). Dependent variables include participant response accuracy, number of missed signals, and mission performance on other tasks in addition to the primary tasks of navigation and mission completion. In addition, participants filled out a series of subjective self-report assessments of the task itself and their engagement in it.

The larger purpose of this study is to produce data on soldier performance under time stress which may eventually be integrated into a larger body of research to develop a dynamic model of stress and soldier performance in combat.

METHODOLOGY

Participants: Participants were male students solicited from undergraduate classes at the University of Central Florida.

Equipment: The experimental facility was equipped with two PC workstations. Each computer system was equipped with a 17" color monitor and driven by a 1.3 GHz processor with 256 MB of RAM. The signal to the participant's monitor was duplicated and transmitted via an S-video cable so that the experimenter was able observe and digitally record the participant's progress through the environment from the second workstation.

The primary platform for this task is a commercially-available video game called Delta Force: Land Warrior, developed by Novalogic inc. This platform was chosen for its face validity and its wide acceptance across various groups involved in the Army MURI (Multi-disciplinary University Research Initiative) project.

Experimental Procedure: In the control condition, participants were requested to perform mission objectives without delineation of consequences for lack of speed in accomplishing said objectives. In the experimental condition, participants were informed that accomplishing the mission objectives would be impossible upon expiration of the time limit and the simulation would end unsuccessfully. Further, they were informed that the time necessary to complete the mission will be very close to the amount of time allotted, and that speed was therefore essential for successful performance. It was expected that participants assigned to this condition would experience significant time-pressure in carrying out the task.



Figure 1: A screen capture from the experimental platform, Delta Force: Landwarrior.

After completing a series of questionnaires, participants were presented with a training scenario in order to orient themselves with the controls used and the flow of the game. The time taken by each participant to complete the training scenario was standardized to ensure that participants had similar training backgrounds. Participants were then asked to repeat the training mission without online feedback from the experimenter. Following this, they performed a simulated hostage rescue mission in hostile territory. During this section of the study, subjects were required to successfully navigate to 7 geographic points (Nav Points) while engaging hostile armed targets. Their objective in this scenario was to successfully rescue as many hostages as possible from a POW camp. Upon completion of this objective, participants were required to escort the hostages to a waiting helicopter for extraction. Participants then completed the rescue mission a second time, without online feedback from the experimenter and with their performance being digitally recorded.

RESULTS

The data presented below fall into two groups: a 13 participant pilot testing run and a 10 participant (to date) experimental run. In addition, only that portion of the data collected by subjective measures is available at this point in the analysis. These include self assessments of performance, task difficulty, and a self-judgment of the time stress they experienced.

Manipulation check:

If our manipulation of the briefing condition was successful, we should expect that the experimental group would experience more time pressure than those in the control group. In order to test this assumption, independent groups t-tests were performed on the data from both the pilot and experimental subject runs.

The mean score of participants in the experimental condition of the pilot group (M = 3.80, SD = 0.837) was not found to be significantly different from that of participants in the control condition of the pilot study (M = 2.67, SD = 1.03),), $t(9) = -1.97$, $p = .080$.

In contrast, the mean subjective rating of time pressure by subjects in the experimental condition of the ongoing study (M = 1.80, SD = 0.447) was found to be significantly greater than that of those in the control group (M = 0.667, SD = 0.5164), $t(9) = -3.844$, $p = .004$, supporting our contention that the manipulation was effective at inducing time pressure.

Perceived task performance: In order to test the hypothesis that subjects in the control group would rate their own performance on the task more highly than those in the experimental group, an independent groups t-test was performed. The mean score of participants in the experimental group (M = 2.20, SD = 0.447) was found to be significantly lower than that of those in the control group (M = 3.00, SD = 0.00), $t(8) = -4.00$, $p = .016$. This indicates that participants who were under less time pressure felt that they performed better on the task.

LESSONS LEARNED

Several issues were addressed during the planning and execution of this study. A discussion of these issues and the methods used to overcome them may prove fruitful to others attempting to perform similar studies in the future.

- When dealing with a video game platform, it is necessary to provide subjects with extensive training in the use of the system prior to data collection. We found that a minimum of 45 minutes or three complete missions was necessary to produce a minimum level of competency in subjects. Failure to provide adequate training may result in a great deal of variability between subjects, making actual trends difficult to identify.

- It is important to recruit participants who are as similar as possible to the population to which the results of the study are intended to generalize to. In the case of a combat study, it might be argued that only male subjects should be used, as women are not currently permitted to serve in combat roles in the United States Armed Forces.
  - This choice has an additional advantage when dealing with "1st person shooter" video games, as we have observed anecdotally that males tend to be more experienced with them and require a shorter training time to asymptote than do otherwise comparable females.

- When manipulating a variable such as time stress by means of a briefing condition, we found anecdotally that exaggerating the strength of the phrasing and tone of the voice to express urgency in the experimental condition resulted in a stronger subjective participant response to the manipulation. In addition, use of a recorded message rather than an in-person briefing appears to standardize the briefing conditions and further reduce variability between subjects.

DISCUSSION

The data produced to date support several conclusions. First, our manipulation check (perceived time pressure) indicates that participants in the experimental condition do feel more pressured than those in the control group. This was an initial concern as the pilot data on this study failed to produce a significant difference as a function of our manipulation. In addition, the data collected to date appear to support the conclusion that participants experiencing time pressure tend to rate their performance more poorly than those who experience no time pressure. This seems to indicate that, at least subjectively, time pressure influences operator perceptions in a simulated combat mission.

REFERENCES

Allan, L.G. (1979) The perception of time. *Perception and Psychophysics, 26,* 340-354.

Allender, L., Lockett, J., Kelley, T.D., Salvi, L., Mitchell, D., Headley, D.B., Promisel, D., Richer, C., & Feng, T. (1995). Verification, validation, and accreditation of a soldier-system modeling tool. *Proceedings of the Human Factors and Ergonomics Society*, **39**, 1219-1223.

Doob, L.W. (1971) *Patterning of Time.* .New Haven: Yale University Press

Fatkin, L.T., King, J.M., & Hudgens, G.A. (1990). Evaluation of stress experienced by Yellowstone Army firefighters (Tech. Mem. No. 9-90). U.S. Army Research Laboratory, Human Research and Engineering Directorate. Aberdeen Proving Ground, MD.

Hancock, P.A., & Desmond, P.A. (Eds.). (2001). *Stress, workload and fatigue.* Lawrence Erlbaum, Mahwah: NJ.

Stanney, K., Salvendi, G. Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *International Journal of Human-Computer Interaction, 10*, 135-187.

ACKNOWLEDGEMENTS

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

TRAINING AND PERFORMANCE OUTCOME: FEATURES OF ADVANCED TECHNOLOGIES MAY NOT BE CONTRIBUTORS

C.S. MORRIS
R.W. TARR
E.C. SHIRKEY
P.A. HANCOCK
M. MOULOUA



ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Introduction.

Technology based training – what features actually influence training and outcome performance.

- Mock-up, motion base, immersion, reality

- Advanced technologies are used to train operators to use their existing or acquired skills in conjunction with other advanced technologies.

Issues : where is advanced technology sufficient?? Necessary??

Introduction.





- While technologically advanced simulators are rapidly being employed as training devices by commercial and defense entities, low cost alternatives are being investigated for more optimal return on investment.

- Recent objectives "Low Cost" – Effective Training Solutions

- Complex interplay of factors (low cost = OTS = mass sales = less specificity and application)

- Only R&D with a human-centric aim

Introduction.

Benefits of low-cost, off-the-shelf PC games:

- Adaptable and portable – can be used in almost any environment
- Synthetic environments that can be linked together for multi-individual and multiple team interactions.
- Mission editor packages

PC based games are being used in 3 ways depending on the quality and appropriateness of the content:

- Classroom enhancement
- Down time skill practice
- Training supplements

Introduction.

-No current formal process or VV&A system for effective employment of alternatives such as these.

-Generally , because of the low cost nature, are implemented as is without outcome performance evaluation.

Background.

Previous PC game based training research:

- Templates

Detection & Perception

- Locating /orienting .903
- Spatial .942
- Perception / Reaction .952

Understanding

- Situation Awareness .690
- Memory .923
- Adaptation .721

Decision Making

-Critical Thinking .887

-Visualization .885

-Problem solving.930

- Mission planning .710

-Mapping .819

-Resource allocation .741

Management & Execution

- Communication .746
- Delegation .885

-Time Management .885

Subject Matter Experts

Objectives.

Does adding supplementary stress to low cost PC based game training environments influence performance?

Ex: Williams, 1980

Can models be derived to explain individual differences in stress enhanced simulation based training and outcome performance.

- Performance is dependent on mastery of stress or adaptation to stimulus induced reactivity

Ex: Caterini, Delhomme, Descheumed-Molinaro, & Dittmar, 1995)

Method.

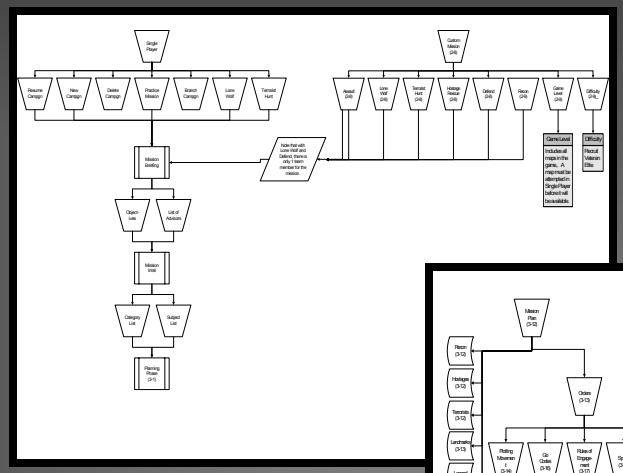
Participants.

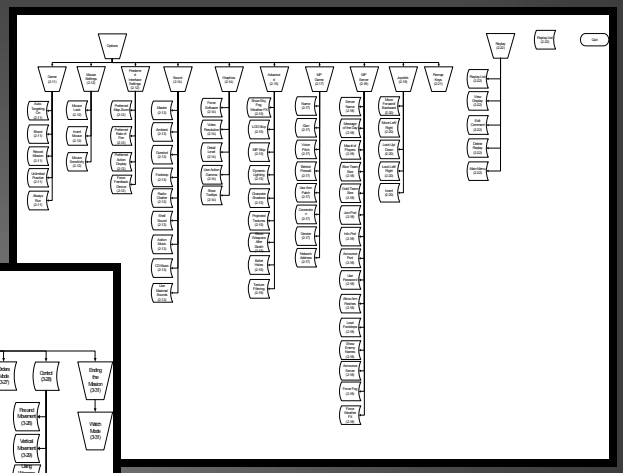
- Twenty-four undergraduate and graduate students.
- Fifteen males and nine females.
- Age (*M* = 22.5, *SD* = 3.65)

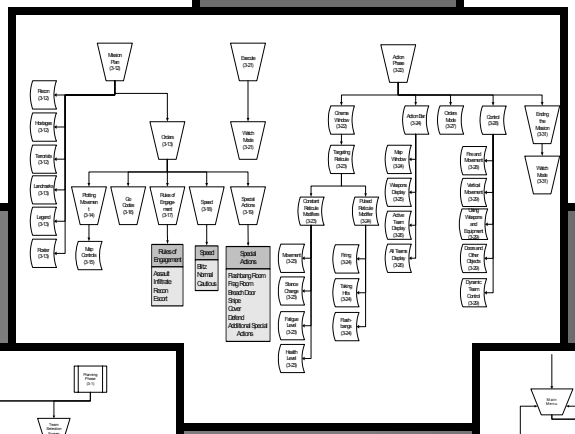
Materials.

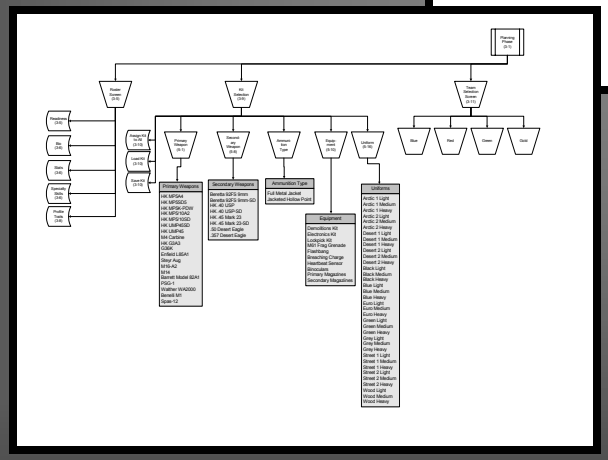
- Delta Force PC based game & supplementary infantry tactic training materials.
- Performance measures based on training objectives.
- Two, fifteen-minute video clips:
    - WWII Documentary (b & w)
    - Saving Private Ryan
- Personality, game habits, pre-post measures of stress

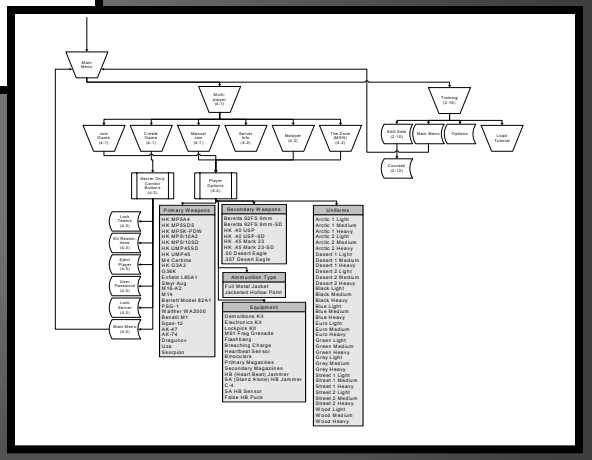
Method.



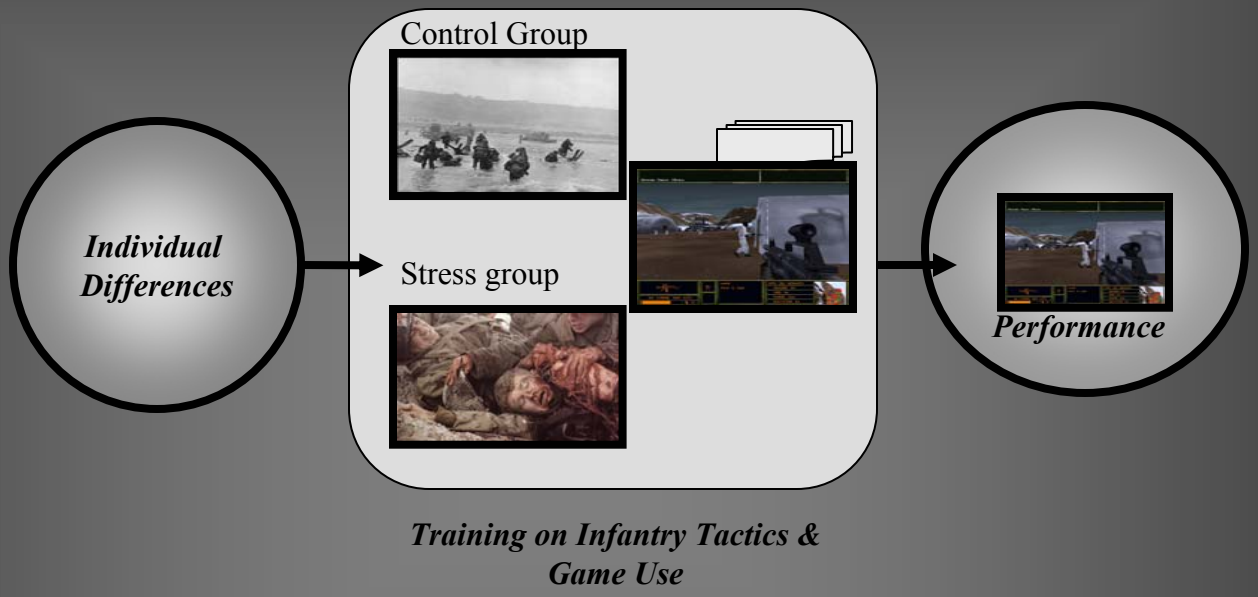








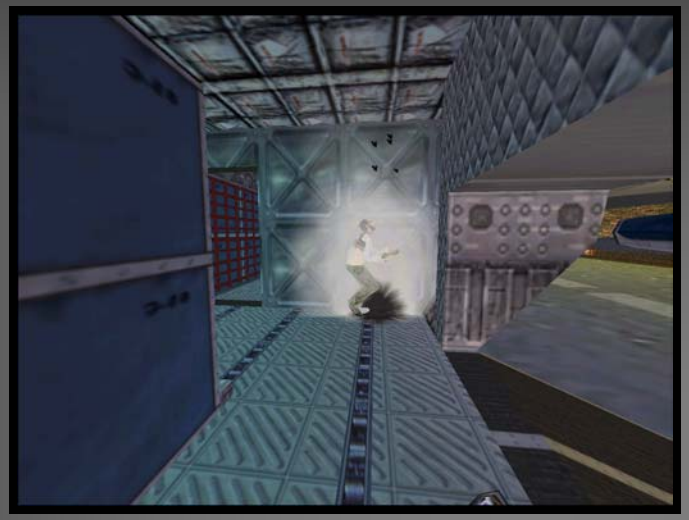
Method.


Control Group
Stress group
Individual Differences
Performance
Training on Infantry Tactics & Game Use

Method.


Eliminate Platoon Officers.
M4 CARBINE, SEMI AUTO
22
Lestat
Wind: N, 2kph
WP6: OBJ MONARCH
HEADING: NA
GRID: NA
RANGE: 42m
ELEVATION: 4m

Method.





Method.





Method.





Method.



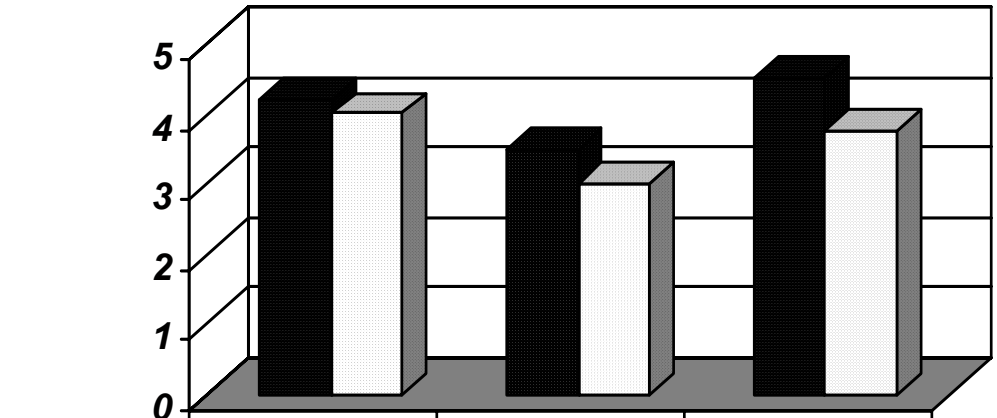


Method.



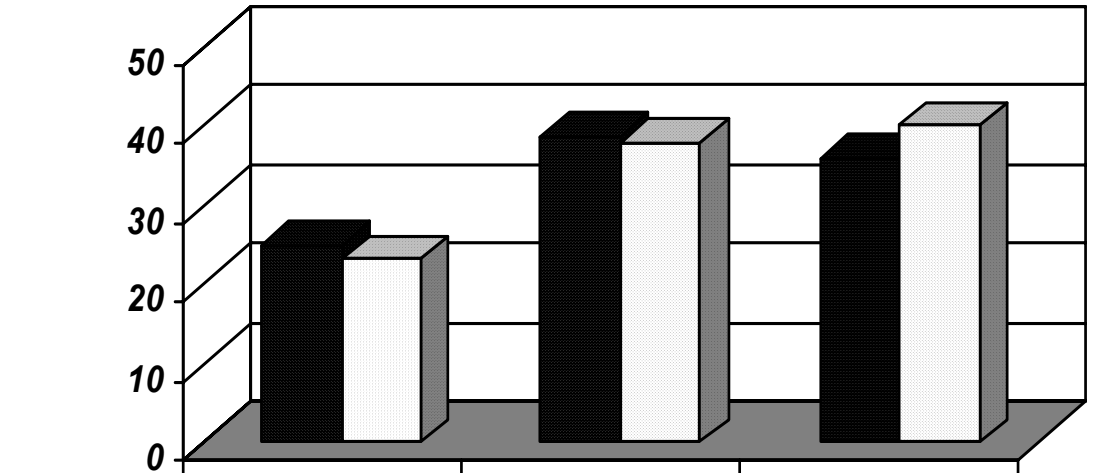


Results.



	Attention to training	Confidence after training	Effort given during mission
Group 1	4.2	3.5	4.5
Group 2	4	3	3.77

Results.



	Movement	Tactics	Equipment use
Group 1	25	38.7	36.1
Group 2	23.22	38.11	40.11

Results.

Disgust $t(22)$ –2.057, p = .05

Anxiety $t(22)$ –2.48, p = .02

Intensity $t(22)$ 2.28, p = .033

Clenched $t(22)$ 2.05, p = .052

Restful $t(22)$ –1.77, p = ..09

R^2 =.62

Gender = -.47

Intensity = .43

Condition = .59

Anger = -.47

Disgust = .55

Discussion.

While enhanced graphics and sound effects increased learners' self reports of stress indicators, intensity, and disgust they did not effect performance outcome.

Implications and theories

IS THERE A SCIENCE OF (ADAPTIVE) INTERFACE DESIGN FOR DECISION-MAKING?

P.A. Hancock
Institute for Simulation and Training, and
Department of Psychology
University of Central Florida

The ability of any force to exert military domination lies in the effective use of its available weaponry. In today's world, such equipment is overwhelmingly computer-controlled. Great effort and expenditure goes into the science behind the conception, the creation, and the testing of such weapon systems. It is therefore more than somewhat disturbing that the interface that provides the critical link between the soldier and the computer that controls the ordnance, has virtually no fundamental scientific foundation for its development. In a charitable way, most contemporary interface designs can be described as the synthesis of tradition-accepted heuristics spread liberally across the contemporary desktop metaphor, tied together with ad hoc programming patches where this marriage of convenience fails. This cannot be said to describe all interface design since there are rare exceptions, such as the ecological approach, that are guided by foundational theory. However, the question addressed in this work is not the value of ad hoc versus principled design strategies per se. Rather, it is to address the more fundamental question as to whether a science of interface design is itself possible. This question is not an empirically resolvable one but rather is an effort based in theory, synthesis, and logic. Failure to resolve this question will doom us to a series of evermore or less 'preferred' designs that vacillate with the times, contingent upon the momentary desires of the individual at hand. Such directionless efforts will cost multiple trillions of dollars in wasted resources and worse, it will result in catastrophic failure at unpredictable but inevitable at crucial operational moments.

We can sharpen the nature of the question of whether a science of interface design is possible by reference to the concept of optimality. Using this term we can pose the more tractable question: is an optimal interface possible? It is one of the more useful characteristics of science to be able to prescribe optimality. Further, to establish optimality, it is important to comprehend the nature of measurement of interface quality. This is true since even if an optimal interface were conceptually feasible, without the ability to measure progress toward or away from such a condition, optimality becomes an empty promise. To begin the exploration we have to acknowledge that optimality in the present case is not a one-time state but a continuously changing target. Given the non-stationarity of human beings we can readily appreciate that they are easily bored and thus any interface that provides the same static structure and stimulation over a long period would rapidly degenerate from optimality (which interestingly is the same argument used by Thomas More against stasis in Utopia as a societal organization). Like More and Godel, among others, we must seek to solve this problem at a meta-level. Meta-level analysis also helps address the problem of task instability since it is likely that as humans are in a state of constant change, the decisions they are required to make relate to task demands that are also in a state of constant flux. What is not in flux are some of the inherent and resident capacities of human decision-makers and it is such stabilities of which we need to take extensive advantage. At present, we cannot claim that an optimal interface can be specified – if indeed it ever can be.

However, recognition of certain inherent problems has brought forward a number of crucial design questions for decision-support interfaces. Primarily, we have evaluated how the task domain should be mapped to the interface characteristics. In past efforts related to decision-support, early designs sought to provide a display for each system feature and a comparable control for what were essentially, system 'main effects.' This still works well in some circumstances such as simple vehicle control. However, the blossoming of system interactive effects soon defeated this design aspiration. Today, we are faced with any number of ways in which task demand can be illustrated on the interface and linked to the selection of effectors to implement decisions made. Such a linkage could be homeomorphic (in somewhat an analogous a manner as the vehicle example given). The present capacity for display space is large and growing but for many of the decision tasks the military are concerned with, homeomorphic mapping is still beyond our technical capacity. A subsidiary case of homeomorphic mapping is isomorphic mapping. Isomorphic mapping requires that specified aspects of the decision space be mapped 1:1 with characteristics of the interface. However, isomorphic mappings do not specify the degree of completeness (i.e., as with homeomorphic mapping where all points are mapped on a 1:1 basis), or the level of mapping (i.e., physical items mapped vs. emergent properties mapped). Thus isomorphic mappings are able to encompass a larger universe of display opportunities but only at the sacrifice of some degree of direct representation. The natural extension is to a polymorphic mapping in which one-to-many and many-to-one linkages are possible. Many current interfaces adopt this form of mapping with no consistent rule about how such mapping is achieved within that interface and certainly no standardization across

different interfaces. At the present time, there appears to be few if any fundamental principles that specify which forms of mapping are suitable to which types of decision (and by extension) operational domain. The design principle we advocate is orthogonal to the design possibilities discussed above and seeks to understand how human attention is directed around different forms of display and to seek to replicate, in the display itself, these attention-focusing aspects of stimulation. We know that intensity, novelty, and complexity draw attention, as does motion and threat in the visual periphery. Further, we know that attentional filters are pre-tuned to specific information and that such tuning can be altered by training. Further aspects of this search for optimal decision-support will be adumbrated.

Part of the present work was prepared through participation in the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009.

Performance Under Stress: A Re-evaluation of a Foundational Law of Psychology

Peter A. Hancock, H.C. Neil Ganey, & James L. Szalma
MIT^2 Lab, Department of Psychology, University of Central Florida
Orlando, Florida, 32816

ABSTRACT

With the intrinsic variation that exists in the behavior of responsive organisms, there are relatively few reliable relationships that have been established between the demand of common elements of an environment and observed reaction. This is compounded in human behavior, in which sentience and the ability to manipulate the environment in a sophisticated manner provides additional possibilities in an information replete ambient display. For these reasons, the few phenomena that have been elevated to the status of a lawful relationship should be treated with some respect, as they provide particular insights into human action by the nature of their ubiquity. In the current work, explanation will be provided about how the work of Yerkes and Dodson (1908) has been misattributed as a law, and what direction the field of stress and performance can go from here.

1. INTRODUCTION

In the relatively short history of research on human behavior, few relationships have been established to such a degree so as to be referred to as 'laws.' The few phenomena that have been elevated to this status are consequently all the more valuable, since they provide unique insights into the activity of humans. Extant laws in psychology appear to share some common threads in that they are founded frequently on transformations of raw data which provide the linear descriptions in Cartesian coordinate space. Fitts' law for the relationship between Movement Time (MT) and the speed and accuracy of normative voluntary movement (Fitts, 1954) is one such law. More recently, the work of Sternberg (1966) has received recognition for its lawful robustness. So, many current 'laws' of behavior are tied to the entropic aspects of information-processing, and thus embrace both descriptive simplicity and theoretical clarity in their exposition.

While the value of such laws should be recognized, they should be subjected to all the more investigation since their effects have a critical impact beyond psychological research alone as they are the guide for design and operational procedures in many complicated technological systems. It becomes even more imperative when both the theoretical foundation and the fundamental description of the law come under question. It is for these reasons that it is essential to reconsider the foundation of one of those laws, the Yerkes-Dodson law. Our purpose here is to re-evaluate the law attributed to them and to explicate its robust perpetuation, in spite of more recent and more sophisticated descriptive and theoretical developments. Originally, the relationship Yerkes and Dodson found was generated to describe the link between discrimination learning and aversive reinforcement. Only after many years was it used to relate the effect of stress on human performance capacity through the introduction of the mediating influence of arousal. Since arousal itself has been the subject of much recent contention (Neiss, 1988) a re-evaluation of the theoretical veracity and descriptive foundation of the Yerkes-Dodson law is clearly mandated.

2. THE DANCING MICE

One misconception that may be perpetuated around the Yerkes-Dodson study is that the experimental work used human participants. Rather, it is the case that following earlier work (Yerkes, 1908) their experiments concerned the behavior of dancing mice. Also, unlike the interpretation that is made in the majority of following studies, the original focus was on learning, not on performance. Specifically, the authors designed three experiments to examine discrimination learning under different conditions of black/white discriminability.

Their experimental set-up is represented in Figure 1, which is reproduced from the original work, illustrating the path of the mice.

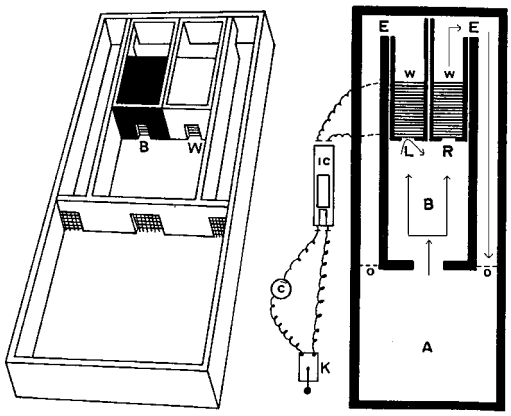


Figure 1: The apparatus utilized by Yerkes & Dodson (1908)

Yerkes and Dodson ran three experiments along these forced choice lines, using a between-mouse design, with the cell number being four mice (2 males, 2 females) in each of the first two experiments but only two mice per cell (1 male, 1 female) in the final critical procedure. Details of these individual experiments are given below.

In Figure 2, the original results of Yerkes and Dodson are reproduced. This shows that in Set 1, three stimulus levels were evaluated. The first was marginally above the threshold level, while the highest condition was noted as extremely disturbing. The middle value remained just that, an unspecified intermediary. The experiments involving sets 2 and 3, utilized five and four levels of shock, respectively. Additionally, the study loses some of its elegance when we realize that the three data sets are based on different stimulus strength values (the strongest shocks used being different for each of the three tasks). This becomes more interesting when we notice that the fifth stimulus level, present in Set 2, is the one that slopes downward, adding to the linearity of the data.

CRITICS

There have, over the years, been a number of critics of Yerkes-Dodson law, although most would agree with us that this is not the fault of the original authors. Each critic seized upon a different aspect as illustrating the weakness of the relationship between what the researchers found and what the work is purported to show. Brown (1965), in focusing on methodological issues, concluded that the 'law' should be buried in silence. One of the most salient points Broadbent (1971) addresses is current (1970's) state of the art experimentation inappropriately dealing with the exceptions to the inverted-U relationship between arousal and performance. Wachtel's (1967) arguments refer to the downward arm of the curve. His disagreement with the past work on the subject of the inverted-U relationship: (p.421) "…attention becomes more and more focused until it reaches a peak, at which point it gradually diffuses (and performance suffers)."

CONCLUSIONS

When taking an objective look at all of the information, one can make inference with certainty. Though Yerkes and Dodson were pioneers in the realm of experimental psychology, they are not the only parties responsible for the theory regarding performance changes as the level of stress is affected. Since many different people in many different areas produced work which has brought us to this explanation, the names Yerkes and Dodson should therefore not be linked with this law. Yes, they did stumble upon the curvilinear function. However, at the time, they were not equipped to confront its implications. The science, in its infancy, was much too young to understand the true depth of the findings. If you look over the studies conducted between Yerkes and Dodson in 1908 and Broadhurst in 1957, there is no indication of curvilinear data, much less discussion about its meaning. To them, it was an anomaly. They were at a loss to explain its true meaning. That responsibility fell to future parties. Just as authorship is given in certain expected manners, so should identification with laws.

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was Supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

REFERENCES

Brown, W.P.I. (1965). The Yerkes-Dodson law repealed. *Psychological Reports*, **17**, 663-666.

Broadhurst, P.L. (1957). Emotionality and the Yerkes-Dodson law. *Journal of Experimental Psychology*, **54**, 345-352.

Fitts, P.M. (1954) The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, **47**, 381-391.

Neiss, R. (1988) Reconceptualizing arousal: Psychobiological states in motor performance. *Psychological Bulletin*, **103**(3), 345-366.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, **153(3736)**, 652-654

Wachtel, P.L. (1967) Conceptions of broad and narrow attention. *Psychological Bulletin*, **68**(6), 417-429.

Yerkes, R.M. (1908) *The Dancing Mice*. New York: The Macmillan Co.

Yerkes, R.M. & Dodson, J.D. (1908) The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, **18**, 459-482.

OFF-THE-SHELF GAMES AND TRAINING: LESSONS LEARNED

C.S. Morris*
Laboratory for Applied Sensory Engagement Research (LASER)
University of Central Florida
Orlando, FL 32816

P.A. Hancock, Ph.D.
University of Central Florida
Institute for Simulation and Training
Orlando, Fl 32816

ABSTRACT

While there appear to be benefits of using low cost, off the shelf game technologies for training (e.g., cost, trainee motivation) there are many drawbacks with respect to doctrinal correctness and negative training transfer. This presentation will describe the various models, guidelines, and validation processes which are being developed for combining and using game elements, segments, scenarios, and practice missions for training.

1. INTRODUCTION

Technologically advanced simulators are rapidly being employed as training devices by commercial and defense entities and low cost alternatives are of interest for a comparable and more optimal return on investment. Notably, low cost off the shelf PC games are being easily acquired and used for classroom enhancement, off-time skills practice, and as training supplements. Recent advances in computer graphics, sound effects, artificial intelligence, networking and teaming capabilities provide more appealing reasons to use the software for training solutions. This form of training is often implemented without consideration of outcome performance or transfer of training. While empirical investigations and sound research exists, the findings are often inconsistent and misleading. A recent focus of the DoD is to determine which of the advanced training technology features and simulation systems actually contribute to an effect on operator training and performance outcome. For example, how necessary is motion base, complete mock-up, large screens, HMD's, versus low cost pc based simulations or off the shelf technologies for influencing immersion, contextual arousal, motivation and thus optimal training and practice conditions. A recent review of military and commercial applications of simulation-based training indicated a severe lack of research data regarding training effectiveness especially on specific transfer of training (Thurman, and Dunlap, 1999).

1.1 Benefits of OTS games.

PC based games are appealing not only for low cost solutions, but also for their adaptability and many uses. For example, embedded mission rehearsal capabilities are available for training anywhere and at anytime. The synthetic environments can be easily linked together for effectively training advanced skills, and shared cognitions in team training further facilitated by mission editor packages that allow for specific scenario and mission construction (Salopek, 1998).

The most notable advances in PC games have come in computer graphics, sound effects and immersive features (Laird, 2000) at a low cost. Additionally, commercial game developers are recently using reputable military data sources for developing their game models thus rendering a doctrinal face validity and making the OTS games even more attractive to the military (Coleman and Johnston, 1999). Likewise, artificial intelligence is representing more human characteristics, suggested to be enough to apply to tactical enemies, partners, support characters, and even units of individuals, such as a platoon (Laird, 2000). Because of this, it was suggested that the single most important feature recommended was currency of game, saying essentially that any game older than 1 year should not be considered in studies.

The motivational quality underlying games is suggested to come from challenge, uncertainty, and complexity. While these features impose a high degree of learner control they can inhibit training and acquisition of basic skills. The motivational qualities of games have been more useful in effecting the amount of time and quality of practice (Driskell, Hughes, and Garris-Reif, 1998).

1.2 Potential drawbacks and considerations

Traditionally, OTS first-person shooter games have been cumbersome and complicated to navigate, setup, execute, and succeed in missions. The games are designed to last for a long time in the hands of expert gamers and even communities that use cheat notes. Consequently, there are many options, missions, levels, scenarios, tools, weapons, and personalities that users must filter through and select. Expert gamers would need to be recruited just to interpret and prepare a scenario for classroom or practice uses and therefore the cost effectiveness is sacrificed. An effective solution to this is to have an expert gamer develop a game map. A game map reveals, in hierarchical form, a graphical presentation of each step of setting up all possible options, scenarios, missions, teams, etc. With this pictorial presentation, the trainer can easily match the training requirements to the potential game materials, options and setups and utilize the OTS environment in the fashion specifically relevant to the training.

When we refer to pc based games involving warlike activity as its objective, the perceived reality that the user experiences regarding the situation and people involved can play a major role in the amount of stress one experiences about the simulated situation (Wilson, Skelly, and Purvis, 1999). The most common criticism of using pc-based games for training, especially in a military environment, is the lack of situational stress and context of war. In a recent study geared at determining the doctrinal and contextual correctness of current pc based infantry games, most Army subject matter experts indicated that the games could not provide the level of realism needed to prepare individuals for tactics in the midst of real war (Morris, and Tarr, 2002). They also indicated that soldiers, if trained using games, would probably not be prepared for the mindset of war because the entertainment condition is drastically different in context (e.g., entertainment verses survival).

Considering the problematic context of entertainment, evidence shows that supplementing stress in game training may effectively produce necessary trainee reactions. Baker, Ware, Spires, and Osborn, (1966) induced threat via a severely displeased military supervisor during simulated combat game. Some individuals were stimulated by the stress and produce more effective performance while others became disorganized and poorly performed. More recently, Morris, Shirkey, Hancock, and Mouloua, (2002) examined the effects of supplementing an infantry OTS game based training session with realistic grotesque war scenes. This intervention produced more motivation and OTS game mission success in individuals than the unstressful control counterpart. It is suggested that contextually-enhanced material that produces positive or motivating stress, can produce higher outcome training performance from OTS games.

REFERENCES

Coleman, D. S., and Johnston, J. H. (1999). Applications of commercial personal computer games to support naval training requirements: Initial guidelines and recommendations. *Proceedings of the 21st Interservice/Industry Training Systems and Education Conference,* Orlando, Florida.

Driskell, Hughes, & Garris-Reif (1998). Personal communication.

Laird, J. E. (2000). An exploration into computer games and computer generated forces. *The eighth conference on computer generated forces and behavior representation.,* Orlando, Florida.

Morris, C.S., & Tarr, R.W. (2002). Templates for selecting PC based synthetic environments for application to human performance enhancement and training. *Proceedings of the IEEE Virtual Reality 2002 conference,* IEEE, Inc.

Morris, C.S., Shirkey, E.C., Hancock, P.A., & Mouloua, M. (2002). The effects of context relevant stress on game-based training. Presented to the American Psychological Association, Chicago, IL.

Salopek, J. J. (1998). Workstation meets playstation. *Training and Development,* 26-35.

Thurman, R. A., and Dunlap, R. D. (1999). Assessing the effectiveness of simulator-based training. *Proceedings of the 21st Interservice/Industry Training Systems and Education Conference*, Orlando, FL.

Wilson, G. P., Skelly, J., & Purvis, B. (1999). Reactions to emergency situations in actual and simulated flight. *Human Performance in Extreme Environments,* **4**(2), 34-45.

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. Dr. P.A. Hancock is the Principal Investigator.

STRESS AND HUMAN INFORMATION PROCESSING: A DESCRIPTIVE FRAMEWORK PRESENTED IN A NOVEL MANNER

P.A. Hancock, P. Ward, J.L. Szalma, S. Stafford, and H.C.N. Ganey
University of Central Florida
Orlando, FL 32816

INTRODUCTION

Understanding the effects of stressors upon human performance and how these effects differ as a function of task, information process and/or mediator has been a long-standing endeavor of the scientific community. Research that has attempted to collectively address some of these issues has examined the relationship between a range of stressors, such as anxiety, heat, and noise, upon task performance assessing, for example, alertness and short term memory, with the aim of revealing common features that apply across tasks (e.g., Hockey & Hamilton, 1983). More recently, research has examined the stress-performance relationship and has identified a number of potential factors which may moderate this relationship, such as social support, locus of control, perceived control, trait anxiety, self-efficacy, self-control, and experience (e.g., Weaver, Morgan, Adkins-Holmes, Hall, 1992; see also Bowers, Weaver, & Morgan, 1996). To date, research on stress and performance has been concerned with specific sources of stress in particular contexts with tasks that require specific forms of information processing. A broad based comprehensive descriptive framework has yet to be formulated. The current work aims to develop such a framework, building upon earlier work by Hockey and Hamilton (1983) and Bowers, Weaver and Morgan (1996).

A difficulty in describing the effects of stress on performance is that such effects depend upon features of the environment (including the task) and of the individual operator. The transactional theory of stress addresses this issue by defining the construct in terms of person-environment interactions. Within the framework of a transactional model, stress states may be viewed as abstracted representations of the relation between individuals and the external demands placed upon them (Matthews, 2001). Lazarus and Folkman (1984) defined psychological stress as a case in which individuals appraise their environment as taxing or exceeding their resources and/or endangering their well being. The negative effects of stress are most likely to occur when individuals view an event as a threat, and when they assess their coping skills as inadequate for handling the stressor.

A second difficulty in creating a descriptive framework is that the effects of various sources of stress are not uniform across all forms of information processing. Indeed, Hockey's (1984) cognitive state model relates particular sources of stress in the environment to specific *patterns* of cognitive activity and performance change. Thus, different environmental stressors are associated with different patterns of change in information processing. Moreover, such changes can be associated with either the structure of these processes (e.g. working memory demands; rate of information transmission) or the strategies individuals employ in response to stress (e.g. allocation of resources; decision criteria; Hockey & Hamilton, 1983). Strategy and structure effects seem to reflect different aspects of cognitive function, and as such, should show different performance outcomes in the presence of particular stressors.

The descriptive framework we propose would incorporate both theoretical perspectives, as well as that of Hancock and Warm (1989), who pointed out that the tasks that operators perform are themselves sources of stress. Our framework defines a three dimensional matrix. Along one dimension will be forms of information processing, including information processing requirements for perceptual and cognitive tasks. The second dimension will be sources of stress (e.g. noise, temperature, social stressors, etc). Finally, the third dimension consists of moderators of stress and performance. These are variables that influence the relation between the sources of stress and the information processing for task performance, and will likely depend to some degree upon the transaction between the operator and the task.

At first glance, this project provides a vast empirical resource from which the effects of different stressors across a variety of information processing tasks can be examined. Moreover, the current program will identify shortfalls in the literature base highlighting potential avenues for future research. At a statistical level, the matrix also serves as the basis for a meta-analytic review providing the impetus for meaningful data to drive theoretical developments and to test existing models of stress and performance.

An objective in the development of the matrix was to inform current evaluatory processes used to model soldier-system performance under stressful conditions and to examine the taxonomic classification (taxons) used in performance prediction within IMPRINT (Improved Performance Research Integration Tool) (see O'Brien, Simon, & Swaminathan, 1992; Allender, Kelley, Archer, & Adkins, 1997). The current system within IMPRINT used for deriving taxons is mainly adapted from Berliner's (1966) task taxonomy with some attempt made to incorporate Wickens' (1981) structure for processing resources (see O'Brien et al., 1992). The matrix will provide both a theoretical and empirical basis for supporting or modifying the existing taxonomy used for task prediction, or for creating a new taxonomy with greater capability for predicting task performance.

SUMMARY

Our framework describes a three dimensional matrix with types of information processing along the first dimension, stressors along the second, and moderators of the stress-performance relationship along the third. The purpose of the matrix is to provide a comprehensive review of the extant literature which will not only identify avenues for future research but will provide empirical support for addressing current models of soldier-system performance under stressful conditions.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Army Research Organization, P.A. Hancock, principal investigator. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

REFERENCES

Allender, L., Kelley, T., Archer, T., and Adkins, R., "The transition and further development of a soldier-system analysis tool," *MANPRINT Quarterly,* V (1), 1997.

Berliner, D.C., Angell, D., and Shearer, J.W., "Behaviors, measures, and instruments for performance evaluation in simulated environments." *Paper presented at a symposium and workshop on the quantification of human performance.* Albuquerque, New Mexico, August 1964.

Bowers, C.A., Weaver, J.L., and Morgan, B.B. Moderating the performance effects of stressors. In J.E. Driskell and E. Salas (Eds.), *Stress and Human Performance.* Mahwah, NJ: Lawrence Erlbaum, pp.163-192. (1996)

Hancock, P.A. and Warm, J.S. "A dynamic model of stress and sustained attention," *Human* Factors, 31(5), pp.519-537, 1989.

Hockey, G.R.J., "Varieties of attentional state: The effects of the environment." In R. Parasuraman and D.R. Davies (Eds.), *Varieties of attention*, New York: Academic Press, pp.449-483, 1984.

Hockey, R. and Hamilton, P. "The cognitive patterning of stress states," In G.R.J. Hockey (Ed.) *Stress and Fatigue in Human Performance*. New York: John Wiley & Sons, pp331-362, 1983.

Lazarus, R., and Folkman, S., "Stress, appraisal, and coping". New York: Springer Verlag.

Matthews, G. "Levels of transaction: A cognitive science framework for operator stress," In P.A. Hancock & P.A. Desmond (Eds.), Stress, workload, and fatigue. Mahwah, NJ: Erlbaum, pp.5-33, 2001.

O'Brien, L.H. Simon, R., and Swaminathan, H., "Development of the personnel-based system evaluation aid (PER-SEVAL) performance shaping functions," *ARI Research Note 92-50, United States Army Research Institute for the Behavioral and Social Sciences*, 1992

Weaver, J.L., Morgan, B.B., Adkin-Holmes, C., and Hall, J.K., "A review of potential moderating factors in the stress-performance relationship". *Technical Report 92-012*, *Naval Training Systems Center,* 1992.

Wickens, C. *Engineering Psychology and Human Performance*. Columbus, OH: Charles E. Merrill Publishing Company, 1984.

THE POINT OF DIMINISHING IMMERSIVE RETURN; IMPLICATIONS FOR SIMULATION-BASED TRAINING

C.S. Morris*
Laboratory for Applied Sensory Engagement Research (LASER)
University of Central Florida
Orlando, FL 32816

H.C.N. Ganey
J. Ross
University of Central Florida
Orlando, FL 32816

P.A. Hancock
University of Central Florida
Institute for Simulation and Training
Orlando, Fl 32816

ABSTRACT

While effective soldier training is enhanced by practice (Salas, & Cannon-Bowers, 2001), stress-exposure (Johnston, & Cannon-Bowers, 1996), and context, the extent to which simulated practice conditions must replicate the operational environment setting been questioned (Morris, Shirkey, Hancock, & Mouloua 2002). All of the features of advanced training technologies (e.g., graphical quality, haptic quality, motion base, surround sound and other aspects noted to foster Immersion) may not be necessary for effective training. The most recent technological advances in environmental fidelity may actually extend beyond the point of diminishing returns for obtaining optimal human reaction. This contention is supported by several empirical evaluations, neurological evidence, and cognitive fundamentals, all of which suggest that humans effortlessly employ a cognitive "immersive-fill" when appropriate conditions are present.

1. INTRODUCTION

1.1 Advanced Technologies and Soldier Training

The effectiveness of the objective force relies heavily on the acceptance of discoveries growing out of supported research. Concurrently, technological advances are occurring at a rapid rate and researchers in the field of training and technology-human interaction are attempting to integrate both the behavioral and technological advances. The progress made in training theory, conditions of training, training delivery strategies, training effectiveness, and transfer is now evident in the multiple models that now exist. It is evident now that not all the characteristics fostering immersion and effective training rely on technology, but more so the strategic incorporation and interactions of context, motivations, content, and cognitive immersive-fill strategies.

1.2 Cognitive fundamentals: Evidence for immersive-fill

In the early 1900's Psychologists such as Wundt and Tichener studied human behavior through experimental manipulations. Along with this process came the need for researchers to develop false realities for the specific purpose of subjecting humans to alternative situations and understanding their finite responses within them. These efforts have resulted in vast knowledge about the control, use, application, and human reactions to various elements that make up altered realities. A common question has been; "at what point can the setting or environmental condition fulfill that of its real world counterpart?" Kantowitz describes the 3 main elements to an experimental situation as setting representativeness (the physical realism or immersive properties), subject (or person) representativeness, and variable representativeness (Kantowitz, 1988). He and many others have demonstrated that "setting representativeness" is overrated and that generalizability of the behavior in the false reality to the real world situation is dependant more-so on the compatibility of psychological processes and not the improvements in realism.

Research has consistently shown that humans naturally enrich and utilize impoverished information. For example, Johannson, (1973), as cited in Chignell, Hancock, & Takeshita, (1999) discussed research on perceived biological movement where in darkness small lights are placed on a humans joints only. Individuals can distinguish exact movements as well as and male verses female movements by seeing a few small lights only. Neurological mechanisms for this include directionally sensitive neurons in area MT, a region in the visual

cortex, which responds identically irrespective of the shape that is moving. A subsystem exists for encoding motion relations in parallel and separate from the subsystem for preprocessing visual form. Similarly, Crick & Koch, (1997), as cited in Chignell et. al., (1999) describe how humans fill-in by making inferences (e.g., viewing the back of person, we know that they have a face). Moreover each facial feature activates different parts of the brain so our perception relies solely on the fill and integration of separate bits and pieces of information. Schacter, (1996), as cited in Chignell et. al., (1999) described that sensory info (smell, sight etc) about specific memories are stored in different regions of the brain and integrated in a central brain structure. Human memory then is constructive, not based on the external environment but on meaningful simple cues to assimilate a percept. Additionally, people tend to be imaginative and they frequently prefer to construct their own image of a story (e.g., a novel) rather than watch a film director's interpretation. The level of fidelity and realism may not be essential for reproducing real world behaviors in the lab likewise, reproducing the real world in the lab may not be necessary for training and practice.

1.2 Comparing technologies: Evidence for immersive-fill

There are relatively few studies that compare the differences between types of technologies and degree of immersion experienced. Botella et al., (1999) found no differences when comparing subject's sense of immersion in two expectedly differing conditions. In one condition, participants performed a task in a high-impact graphics workstation with a high quality head mounted display (HMD). The other condition used a low graphics/quality HMD with 2-D joystick. Results indicated no statistical differences on dependant variables of presence or reality judgment between the two conditions. From the field of clinical psychology, Weiderhold, & Wiederhold (2000) investigated the efficacy of Virtual Reality experience for overcoming phobia versus imaginal therapy (using one's imagination only). Unexpectedly, the subjects who used the advanced virtual reality system did not differ on physiological indices of immersion and treatment efficacy from those who used imagery to produce the experience of flying. Other findings from Weiderhold, & Wiederhold, (2000) revealed physiological changes associated with level of self-reported immersion. Based on this some would argue that advances in system features would cause greater levels of immersion. We suggest alternatively that the result represents an individual's immersive-fill (a function of motivation, interest, directed energy, and context).

With respect to US military soldiers, it is expected that their mentalistic requirements for immersion are intact. For those who lack motivation to train, alternative contextual strategies have been shown to enhance immersion such as the element of "game" which fosters motivation to succeed and can be relatively easy to facilitate as well as inexpensive (e.g. Morris, Hancock, Shirkey, & Mouloua, 2002).

In sum, while advanced simulations may "aid" in the process of human immersion, the variance associated with degree of immersion has repeatedly been shown to be predominantly a function of individual responsiveness to cues and characteristics of the environment, not associated with fidelity or a replicated reality. As technologies advance so too does the human's reliance on them. A strategic facet of future soldiers will be their ability to adapt to novel situations with minimal information and naturally apply context for effective performance.

REFERENCES

Botella, C., Rey, A., Perpina, C. Banos, R., Alcaniz, M., Garcia-Palacios, A., Villa, H., & Alozano, J. (1999). Differences on presence and reality judgment using high impact workstation and a PC workstation. *Cyberpsychology & Behavior,* **2**, 49-52.

Chignell, M., Hancock, P.A., & Takeshita, H. (1999). Human-computer interaction: The psychology of augmented human behavior. In P.A. Hancock, (Ed) *Human performance and ergonomics* (pp 291-328), Academic Press Inc.

Johnston, J.H., & Cannon-Bowers, J. (1996). Training for stress exposure. In J.E. Driskell and E. Salas (Eds). *Stress and human performance*. (pp 223-256). Mahwah, NJ: Erlbaum

Kantowitz, B. H. (1992). Selecting measures for human factors research. *Human Factors,* 34(4), 387-398.

Morris, C.S., Shirkey, E.C., Hancock, P.A., & Mouloua, M. (2002). The effects of context relevant stress on game-based training. Presented to the American Psychological Association, Chicago, IL.

Salas, E. & Cannon-Bowers, J.A. (2000). The anatomy of team training. In S. Tobias, & J.D. Fletcher (Eds.).Training and Retraining: A Handbook for Business, Industry, Government, and the Military. (pp 312-335). Macmillan Reference, USA.

Weiderhold, B.K., & Weiderhold, M.D. (2000). *Lessons learned from 600 virtual reality sessions.* Cyberpsychology & Behavior, **3**, 393-400.

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. Dr. P.A. Hancock is the Principal Investigator.

Stress and Human Information Processing: A Descriptive Framework Presented in a Novel Manner

Peter A. Hancock
Paul Ward
James L. Szalma
Shawn Stafford
H.C. Neil Ganey

Minds In Technology MIT² Machines In Thought
INSTITUTE for SIMULATION & TRAINING
University of Central Florida
Multi-Disciplinary University Research Initiative
MURI-OPUS
Operator Performance Under Stress
www.mit.ucf.edu

Abstract

A three dimensional matrix is proposed to detail the interactive effects of stress, human information processing and performance moderators. Our aim is to inform current modeling practice used to predict soldier-system performance. Within IMPRINT, performance is modeled under standardized conditions using the ACT-R cognitive architecture and then modified via the application of degradation algorithms to reflect individual task performance under stress. However, IMPRINT is limited by the availability of suitable degradation algorithms and by application of the same algorithm to groups of task which share similar characteristics (e.g., taxons). The matrix provides resolution to these issues by firstly, educing the divergent stress effects on perceptual, cognitive, and action components of real-world tasks, and secondly, by proposing enhanced degradation factors that will facilitate future modeling endeavors .

Integrated modeling in IMPRINT and ACT-R

IMPRINT is a task network architecture used to represent soldier-system performance for the purposes of procurement, test and evaluation. One of its main aims is to predict performance under a myriad of conditions (Allender, Salvi, & Promisel, 1997). System missions are subdivided into functions and tasks, and the macro-components of real world performance are modeled using Microsaint. IMPRINT has recently been integrated with ACT-R (/PM) (see Anderson & Lebiere, 1998), a hybrid cognitive modeling architecture, to model the micro-components of situated tasks (Lebiere, et al., 2002) (see Figure 2). These models communicate by sending task-related information and the resultant modeled data output over a component object model (COM) link. One of the benefits of using this type of integration is the reduced need for endless data. Moreover, ACT-R can generate distributions of time and error data for each MOS task that appropriately reflects the actual perceptual, cognitive, and motor components of tasks experienced in the real world, albeit under standardized conditions.

Figure 1. Stress degradation algorithms available for each taxon (T = time data only, A = accuracy data only; adapted from Allender, Salvi, & Promisel, 1997)

Taxon \ Stressor	MOPP	Heat	Cold	Noise	Sleepless Hours
Visual	T	A	T		
Numerical		A			TA
Cognitive		A			TA
Fine Motor Discrete	T	A	T		
Fine Motor Continuous					
Gross Motor Light	T		T		
Gross Motor Heavy					
Communication (Read & Write)		A			
Communication (Oral)	T	A		A	

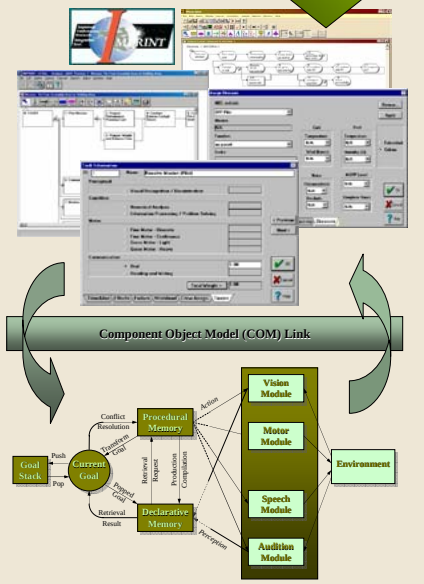


Figure 2. Illustration of the integration between IMPRINT and ACT-R / PM (adapted from Anderson & Lebiere, 1998; Lebiere et al., 2002; Byrne, 2001)

Stressors within IMPRINT

Performance is highly variable under stressful compared to standardized conditions, and some individuals may be more suited to a particular task than others. Consequently, performance shaping functions and stress degradation algorithms are applied to modeled data within IMPRINT, to predict relative change in performance for a specific set of personnel characteristics and training regimen, and to modify typical/baseline measures to account for the negative affect upon performance. Application of each algorithm is determined by task classification (e.g., taxons) assignment so that the same algorithm is applied to tasks that fall within the same taxon. Nine taxons have been identified spanning perceptual, cognitive, psychomotor and communication modalities (see Figure 1). Due to the differential effects of stressors on each task, tasks are weighted according to the taxon to which they have been assigned (Allender, Salvi, & Promisel, 1997). At present, five, predominantly environmental, stressor algorithms exist within IMPRINT. These relate to Mission Oriented Protective Posture gear for biological, chemical and nuclear defense, heat, cold, noise, and sustained operation. However, much of the research used to derive these algorithms is dated and the current depiction does not capture the state of the art with regards to known stressor effects. Moreover, information for each of the five stressors is not available for all taxons and a comprehensive evaluation of current stressors has not been implemented.

The Matrix

To date, very few assessments have been made on a broad range of interactive stressors with the specific intention of detailing both the variable and predictable effects upon perception, cognition, and action. Hockey and Hamilton's (1983) research on the cognitive patterning of stress states indicated that although some commonality exists between particular stressors, a multidimensional approach is necessary to explain the lack of response uniformity. Those variables that moderate the relationship between sources of stress and the nature of task-related information processing are likely to be influenced by the transaction between operator and task (Lazarus & Folkman, 1984). Clearly, the current stressors used within IMPRINT are not exhaustive and other environmental, social, and task-based stressors need to be incorporated within the stress dimension to make it comprehensive. We propose a descriptive framework which details the effects of a broad range of stressors on varied human information processing tasks and examines the potential variables that may moderate this interaction. Collectively, the matrix provides an integrative database that will drive theoretical development, facilitate modeling efforts, and allow existing models of stress and performance (e.g., Hancock & Warm, 1989) to be tested. Empirical verification of modeled cognitive processes will facilitate integration of stress-related effects into future modeling ventures. The integration of new stressors, validation and modification of old stressor effects, and revisions to the existing taxon scheme, once integrated, are hypothesized to all provide an immediate effect upon soldier-system performance prediction within IMPRINT.

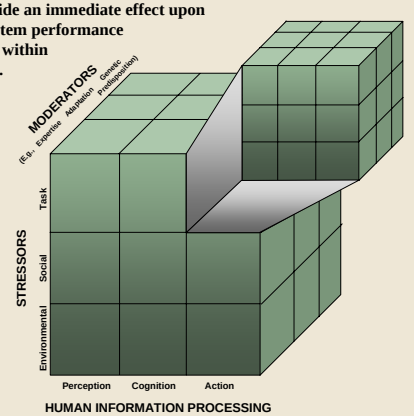


Figure 3. Pictorial representation of the three-dimensional matrix for stress, human information processing and moderators of the stress-performance relationship

ERGONOMICS, 2002, VOL. 45, NO. 14, 991–994

Taylor & Francis
Taylor & Francis Group

Commentary

Sans subjectivity — ergonomics is engineering

P. A. HANCOCK†*, J. L. WEAVER† and RAJA PARASURAMAN‡

†Department of Psychology, University of Central Florida, University of Central Florida, Orlando, FL 32816, USA

‡Cognitive Science Laboratory, The Catholic University of America, Washington, DC 20064, USA

1. Introduction

We much admire and agree with many of the points and issues that Annett has raised and believe his work has great value for both researchers and practitioners in ergonomics and allied fields. We also believe that Annett's paper will join others in giving the lie to the frequent assertion that ergonomics is an atheoretical field, an 'appliance science' (Hancock 1997), not fit for the high company of theoretical psychology and at best a poor cousin to higher forms of engineering. Annett's rich theoretical observations, in contrast, begin to bring to the fore crucial philosophical questions that underpin the ergonomics enterprise.

Somewhat ironically, Annett highlights the central role of theory in ergonomics by discussing the use of subjective measures, which have often been criticized as being fundamentally flawed and *unscientific*, indeed even by some of the founders of the discipline, as Annett points out. We strongly agree with Annett's main points that all empirical observations have a subjective component, and that subjective methods, when developed and applied appropriately in accordance with accepted scientific criteria, have an important place in ergonomics. We raise two additional issues, however, where we believe Annett has not gone far enough and as commentators we solicit his further insight into these matters.

Is the attempt to provide scientific assessment of private experience a categorical mistake?

In the early decades of the 20th century, as Psychology was engaged in the desperate struggle for acceptance into the family of the so-called legitimate or 'hard' sciences, John Watson (1913) called for the suspension of any reference to internal mental states and to restrict psychological discourse to events that were externally observable. Nominally made in response to the mire of anchorless introspectionism, this radical behaviourist reformation excised from psychology the very pith and heart of conscious experience. Today, we still reel from this abdication. The distaste for subjective experience still rankles and periodically the argument resurrects itself in various guises. Here we see Annett engage once again in this struggle, expressed in the terms of ergonomics research. Like many of his forebears, Annett raises crucial ques-

*Author for correspondence. e-mail: phancock@pegasus.cc.ucf.edu

Ergonomics ISSN 0014-0139 print/ISSN 1366-5847 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0014013021016689 8

tions about how to expose private experience to public scrutiny such that acceptable 'intersubjective' levels of agreement are reached about the phenomena at hand. However, this very first step itself may be a categorical mistake. Consider, for example, the following line of argument. It may simply be a mistake to attempt the measurement of mind using the methods created to measure matter. If mind and matter are truly disparate in their very essence, the approach to understanding the one may very well be inappropriate for the other. If we accept the division of mind and matter, it may well be that we have to forego the envied appellation of 'science' and, rather, seek other but nevertheless mutually acceptable methods of evaluation to discover truths about the essential nature of mind. What those methods might be are difficult to conceptualize, but we should not be disheartened since science itself, in the form we know it, is barely five centuries old (Dee 1551, Bacon 1620).

There is, of course, a complete antithesis to the above proposition which is that mind is nothing more than the outcome of a specific arrangement of matter. This modern restatement of a materialist philosophy of mind argues that mental states 'emerge' when neural networks of a sufficient level of complexity and connectivity are activated. The revolutionary impact of cognitive neuroscience on psychology over the past decade is indicative of a similar influence in the years to come on other behavioural and social sciences, including ergonomics, so much so that we have endorsed the call for the development of the field of 'Neuroergonomics' (Parasuraman 2003, Hancock and Parasuraman 2001). Given the advances in today's sophisticated brain-imaging techniques, it may well be that in the foreseeable future we will be able to specify the neuro-physiological patterns that denote specified mental states. The cognitive neuroscience enterprise has already yielded dividends in our understanding of such mental constructs as *attention* and *consciousness*. For example, over a century ago the great German scientist Herman von Helmholtz (1894) showed that while pointing his eyes straight ahead at a single letter projected briefly to a point on a wall in front of him, he could nevertheless attend to and be aware of letters presented at other locations far removed from the fixation point. The subjective sense of watching out of the 'corner of one's eyes' is well known to anyone who has walked warily at night in a crime-ridden neighbourhood or to an animal drinking at a watering hole that is also frequented by predators. Modern cognitive neuroscience research has convincingly demonstrated that this subjective state of visual attention is associated with activation of a neural network involving specific cortical and subcortical brain structures. This in turn has led to improved theories of visual attention and stimulated new and far-reaching kinds of research (for a review, see Parasuraman 1998). Neural measures do not necessarily have primacy over performance measures or verbal reports in this kind of research, and the different measures may not always be correlated with one another. Rather, each type of measure has its uses. In some cases, the *softer* verbal report may even be *superior* to the *harder* neural measure. Consider the case of a stroke patient who as a result of damage to a component of the neural network cannot spontaneously attend to a particular portion of the visual field. While looking straight ahead (as in von Helmholtz's experiment), the patient cannot verbally report, nor is aware of, letters presented to their left while letters are shown to their right —

the neurological condition of neglect. In this case, the subjective state (and verbal report, or lack thereof) is a central feature of the condition at hand and is as important to the clinician attempting to diagnose neglect as is the neural measure to the scientist interested in a theory of neglect and attention.

It may well be that other multidimensional constructs such as fatigue, which have defeated our efforts at purely psychological definition (Muscio 1921, Broadbent 1979, Hancock and Verwey 1997), also turn out to be represented by specific neurophysiological activation patterns which change in understandable ways with variables such as task factors, the environment, etc., in short the global variable called 'The challenge' by Annett in his figure 2. This is not to say that psychological constructs will be reduced to neural measures, for a materialist philosophy of mind is not necessarily also a reductionist philosophy. Rather, theories of psychological constructs will be enriched through a better understanding of the neural systems that implement mental function. Further progress in neuroergonomics will therefore blur the distinction between so-called objective physiological measures and subjective measures. Furthermore, if a particular subjective state is a particular neural state, then it makes no sense to argue that one is superior to the other. Rather, measures of each must be used in the development of a comprehensive theory that can validly be applied in different situations.

This is essentially also Annett's position about the use of subjective methods in ergonomics. However, one's position on the materialist/non-materialist dichotomy predicates all subsequent discourse on the use and scientific status of subjective measures vis-à-vis so-called objective measures. We, as commentators, would liked to have seen Annett's explicit evaluation of these polar extremes. While not expecting him to resolve the mind–body problem, understanding his foundational perspective on a philosophy of mind would have helped frame the observations to come.

2. What is mind?—No matter. What is matter?—Never mind

(We have traced the origin of this phrase to T. H. Key, once Headmaster of University College School, on the authority of F. J. Furnivall (Bartlett 1919). The fact that it is probably best known as a pronouncement of Homer Simpson is a further tragedy of our times.)

Annett's commentary focuses on the problems of subjectivity. However, rather than lamenting the difficulties of this enterprise we embrace them, for without the necessary conscious experience of the individual involved, ergonomics devolves to engineering. Thus, while we do not deny the difficulties of assessment, we do note that, *sans* subjectivity the very notion of ergonomics itself is barren. Cloaked in the guise of these modern concerns of ergonomics applications, Annett touches on the fundamental question of philosophy concerning the nature of what can be known and the nature of that consciousness which can know it. This aspect of his work draws our highest praise because, at heart, ergonomics is concerned with the expression of intention as mediated by technical systems (Hancock 1997).

But Annett does not go far enough. If all measurement is subjective in nature, is not all experience subjective also? We believe that Annett navigates between the Scylla and Charybdis of such questions knowing that solipsism can appear initially to be a very slippery slope. We can be fairly certain there is no empirical resolution to the

proposition that Berkeley (1710/1974) set forth concerning the superfluous nature of matter. Not excepting some of the more delightful obfuscations of certain German philosophers, potential resolution of this issue on the nature of what is real can only be captured conjecturally. But let us not be so afraid of this putative Pandora's box! The eidetic reassurances of the American Pragmatic School have too long held sway. We have come to accept as given that objects and events occur 'out there' in a real world and have parsed consciousness and matter to accord with this notion. Only in our more outlandish movies do we now question the correctness of this assumption. While the foregoing are largely cryptic remarks founded upon the evolution of philosophical thinking, they are directly relevant to ergonomics in the modern world. We see ergonomics as intention in action. That action is mediated by the technologies we have built. If this is an illusion shared by collective minds, the locus of intention is still within the human being and the further exploration of affective states of mind is consequently critical. If the world is an illusion of a single mind, being either personal or deistic in foundation, the way that mind explores itself via the creation of external representations is itself also of value. Rejecting the comfort of pragmatism does not mean that even radical solipsism necessarily devolves to futilitarianism. Actually, such a proposition concerning the locus of intention has much to say about the way in which any reality (including a technologically replete one) can be organized. We applaud Annett's work on several levels. That it opens further discussion and consideration of the very foundations upon which our reality is based attests to the centrality of ergonomics to the human enterprise.

Acknowledgement

This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI), Operator Performance Under Stress (OPUS) Grant DAAD19-01-1-0621, P.A. Hancock Principal Investigator. The views expressed here are those of the authors and do not necessarily represent those of the named agency.

References

BACON, F. 1620, *Novum Organum*; repr. J. Devey (New York: Collier, 1902).

BARTLETT, J. 1919, *Familiar Quotations*, 10th edn (Boston: Little Brown).

BERKELEY, G. 1710/1974, A treatise concerning the principles of human knowledge, in *The Empiricists* (New York: Anchor Press).

BROADBENT, D. E. 1979, Is a fatigue test now possible?, *Ergonomics*, **22,** 1277–1290.

DEE, J. 1551, *Mathematical Preface* (London: Mortlake).

HANCOCK, P. A. 1997, *Essays on the Future of Human–Machine Systems* (Eden Prairie: Banta).

HANCOCK, P. A. and PARASURAMAN, R. 2002, Human factors and ergonomics, in *Handbook of Cognitive Science* (in press).

HANCOCK, P. A. and VERWEY, W. B. 1997, Fatigue, workload and adaptive driver systems, *Accident Analysis and Prevention*, **29,** 495–506.

MUSCIO, B. 1921, Is a fatigue test possible?, *British Journal of Psychology*, **12,** 31–46.

PARASURAMAN, R. 1998, *The Attentive Brain* (Cambridge, MA: MIT Press).

PARASURAMAN, R. 2003, Neuroergonomics: Research and practice, *Theorectical Issues in Ergonomic Science,* **4,** 4–19.

VON HELMHOLTZ, H. 1894, *Handbuch der Physiologischen Optik* (Leipzig: Voss).

WATSON, J. B. 1913, Psychology as a behaviorist views it, *Psychological Review*, **20,** 158–177.

WORKLOAD AND STRESS OF CONFIGURAL DISPLAYS IN VIGILANCE TASKS

James L. Szalma
University of Central Florida
Orlando, FL.

The workload and stress associated with configural displays in two vigilance tasks were investigated. Two kinds of configural displays were employed: A bar graph display and an object display. A non-configural bar graph display served as a control group. Relative to the non-configural display, both configural displays improved performance in a task requiring integration of information, but were not significantly different from the control group in a task requiring focused attention to display elements. The object display reduced workload in both tasks, but the bar graph configural display did not. Results showed a complex pattern of association/dissociation of workload with performance. Self reports of stress revealed that the tasks were stressful but that configural displays did not reduce the stress of either task.

Configural displays containing emergent features have been investigated for their utility in tasks requiring either integration of information from multiple sources or focused attention on specific display elements. These displays have been shown to improve performance on integration tasks but not on tasks requiring focused attention (focused tasks), although there does not appear to be a cost associated with focused tasks (see Bennett & Flach, 1992 for a review). The superiority of configural displays in supporting performance on integration tasks likely results from the presence of emergent features that are well-mapped to system dynamics (Bennett & Flach, 1992).

The effects of configural displays on performance in monitoring tasks have been well explored, but to date, investigation of the workload and stress associated with using these displays has been neglected. If designing emergent features into a display improves task performance by reducing the resources required for information processing, one might also expect such displays to reduce the perceived workload and stress of these tasks relative to a task employing a display that does not contain an emergent feature. This effect is more likely to emerge in a task requiring integration of information rather than in one requiring focused attention to specific display elements. A purpose for the present study was to test the possibility that configural displays might reduce perceived workload and stress in an integration task but leave them unchanged in a task requiring focused attention.

Although workload and stress have not been investigated in regard to configural displays, they have been examined in experiments employing the vigilance paradigm. Several investigations have shown that monitoring tasks that require sustained attention impose high workloads on operators and that they find these tasks to be stressful. In addition, several investigations have shown that display characteristics that impair vigilance performance also elevate the workload and stress of vigilance; factors that improve performance reduce the workload and stress of these tasks (see Warm, 1993 and Warm, Dember, & Hancock, 1996 for reviews).

While performance, workload, and stress have been investigated extensively in vigilance, the use of configural displays to reduce workload and stress has not been examined in this context. In terms of performance, note that most studies investigating configural displays analyzed performance scores averaged across the experimental session. In contrast, vigilance experiments typically analyze performance change over time. Thus, in addition to examining workload and stress, investigation of configural displays within the vigilance paradigm permitted the assessment of the effect of configurality on the vigilance decrement, or the decline in sensitivity commonly observed over time (See, Howe, Warm, & Dember, 1995). Given that increasing signal salience improves vigilance performance (Warm & Jerison, 1984), it was expected that within the integration task the configural displays would attenuate the performance decrement relative to a non-configural display. No such differences were expected in the focused task.

METHOD

Forty eight men and forty-eight women participated in the study, ranging in age from 18 to 46 years, with a mean of 20.8 years. Two levels of task-type (focused attention vs. integration) were factorially combined with three display types (bar graph non-configural, bar graph configural, and object configural) to yield six experimental groups. Sixteen observers were assigned at random to one of the six conditions with the restriction that the groups were equated for participant sex.

Perceived workload was measured using the NASA Task Load Index (TLX; Hart & Staveland, 1988), a multi-dimensional instrument that provides a reliable index of overall workload and also identifies the relative contributions of six sources of workload. Stress state was measured using the Dundee Stress State Questionnaire (DSSQ; Matthews, Campbell, & Falconer, 2001; Matthews et al., 1999) which consists of 11 factor-analytically determined scales and 3 secondary factors that reflect multiple dimensions of cognitive state associated with task performance. The secondary factors are Task Engagement, reflected by the Energetic Arousal, Concentration, and two Motivation scales (Success and Intrinsic); Distress, reflected by Tense Arousal, Hedonic Tone, and Control and Confidence; and Worry, comprised of Self-Focused Attention, Self-Esteem, Task-Related Cognitive

Interference (thoughts concerning task performance), and Task-Irrelevant Cognitive Interference (thoughts unrelated to task performance).

The displays employed in this study were adapted from those used by Buttigieg and Sanderson (1991) and Sanderson et al. (1989), and are shown in Figure 1. Both tasks required observers to monitor the state of an imaginary system consisting of two inputs and an output. In the integration task neutral events, requiring no response from the observer, were cases in which the output of the system was the average of the two inputs. The input and output values in the bar graph displays were represented by the heights of the rectangles. Neutral events for the rectangle displays had two different sets of values: 1) input 1 and input 2 lengths were 5.8 cm and 2.8 cm, respectively, and the output length was 4.3 cm; and 2) input 1 and input 2 lengths were 4.3 cm and 1.4 cm, respectively, and the output value was 3.0 cm. Critical signals for detection were cases in which the value of the output was 1 cm higher or lower than the average of the two input values. The polygon was designed so that the input and output values, represented as the distance from the base to the top of the polygon at the point below each dot, were equivalent to those of the rectangle displays. In both configural displays the emergent feature was linearity, either in the alignment of the input and output bars (bar graph displays) or of the top of the polygon (object display). Note that participants in the integration task were informed of the system dynamics but were not informed of the emergent feature that could be used to aid performance on that task.

In the focused task the input and output values were represented as the vertical distance of each .6 cm diameter black dot from its respective rectangle (bar graph displays) or from the polygon (object display). Neutral events were defined as cases in which each dot was a standard distance (.7 cm) from its rectangle/polygon. Critical signals for detection were cases in which any one of the dots was .5 cm closer than the standard distance to the rectangle/polygon.

Pilot work established that the integration and focused tasks were psychophysically equated for difficulty. In that study, a sample of sixteen observers discriminated critical from neutral events for each task and display format using a two-alternative forced-choice procedure,

In all conditions, stimuli were presented at a rate of 26 events/min, and twelve critical signals appeared during each of the four six minute periods of watch. Observers responded to critical signals by pressing the spacebar on a computer keyboard. Responses occurring within 1.5 sec after the onset of a critical signal were recorded as correct detections. All other responses were recorded as false alarms.

Participants completed the pre-version of the DSSQ after the task instructions and then experienced a 12-min practice vigil. During practice observers received feedback regarding correct detections, false alarms, and misses. The practice session was followed immediately by a 24-min vigil divided into four continuous 6-min periods of watch. To be retained in the study observers were required to detect 70% of critical signals and commit no more than 10% false alarms during the first period of watch. Twenty-seven participants failed to meet criterion and were replaced. Of those participants, sixteen were in the non-configural/integration task group. However, t-tests showed no significant differences in workload or stress scores between observers who did not pass criterion and those who passed. Hence, for the purposes of this study, the two groups differed significantly only in their ability to pass criterion. Note that in terms of performance differences, omitting the poor-performing observers from the non-configural control group actually reduced the hypothesized effects.

After the vigil, participants completed a computerized version of the TLX and a post-task version of the DSSQ. The order in which these measures were administered was counterbalanced to control for potential order effects.

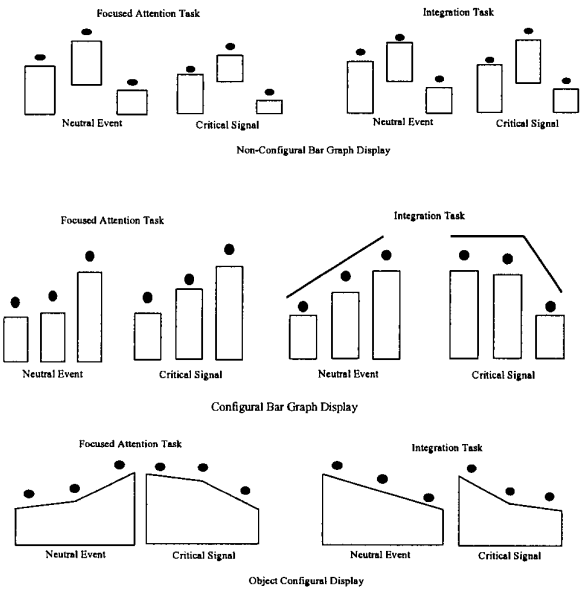


Figure 1. Displays employed in the present experiment. In each case the output is in the middle, flanked by the two inputs. The lines above the rectangles in the configural bar graph display indicate the emergent feature in the display. Note. The labels and the lines above the rectangles are for illustrative purposes and were not present during the task.

RESULTS

Percentages of correct detections and false alarms for each observer were used to compute signal detection theory measures of A' and B_D''. Mean A' scores for each display are plotted as a function of periods of watch in Figure 2. An ANOVA revealed that the only significant main effect was for periods of watch, with sensitivity declining over the course of the vigil, $F(2,205)=18.05$, $p<.001$. However, there was a significant task x display interaction, $F(2,90)=4.71$, $p<.05$, and a significant 3-way interaction between task, display, and periods, $F(5,206)=5.35$, $p<.001$. Note that for all analyses involving repeated measures, Box's epsilon was used to compute degrees of freedom to correct for violations of the sphericity assumption (Maxwell & Delaney, 1990). Separate display x periods ANOVAS within each task revealed that in the integration task there were significant main effects for

periods, F(2,107)=7.00, p<.01, and for display, F(2,45)=5.52, p<.01.

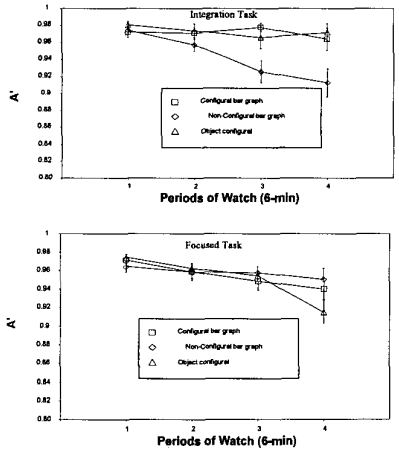


Figure 2. Sensitivity as a function of periods of watch for three display types. Data for the two tasks are presented separately. Note. Error bars are standard errors.

Post-hoc comparisons (the Bonferroni correction was used for all post-hoc tests) indicated that the sensitivity of observers in the two configural display conditions did not significantly differ from one another but both were significantly more sensitive than observers who experienced the non-configural display. There was also a significant display x periods interaction, F(5,107)=4.08, p<.01. Tests for the simple effects periods within each display indicted that sensitivity in the two configural displays remained stable over time (F<1 in both cases), while sensitivity of observers in the non-configural display group declined with time on watch, F(2,31)=10.08, p<.001. In the focused attention task the only significant effect was for periods, such that sensitivity declined over time, F(2,96)= 11.81, p<.001.

B_D scores for observers in each condition and period of watch are shown in Table 1.

Table 1. Means and standard errors for B_D'' for each condition and period of watch.

	INT			FOC			*M*
P	nc	cb	co	nc	cb	co	
1	.55	.80	.83	.75	.77	.73	.74
	(.10)	(.05)	(.04)	(.06)	(.05)	(.09)	
2	.78	.86	.86	.84	.86	.85	.84
	(.07)	(.04)	(.02)	(.04)	(.04)	(.05)	
3	.89	.85	.85	.83	.89	.86	.86
	(.04)	(.02)	(.02)	(.06)	(.03)	(.06)	
4	.88	.83	.81	.90	.86	.95	.87
	(.03)	(.04)	.(04)	(.02)	(.06)	(.01)	
M	.78	.84	.84	.83	.85	.85	

Note. P= Period of watch; INT=Integration Task; FOC=focused task; nc=non-configural bar graph display; cb=configural bar graph display; co= configural object display. Standard errors are in parentheses.

An ANOVA indicated a main effect for periods, such that observers became more conservative with time on watch, F (2,198) = 14.7, p<.01, and a significant 3-way interaction between task, display, and periods of watch, F (4,198)=3.4, p<.01. All other sources of variance failed to reach significance. Separate display x period ANOVAs were calculated within each task. In the integration task significant effects were obtained for periods of watch, F (2,76) = 6.4, p<.01, and the display x periods interaction, F (3,76) = 4.3, p<.01. Tests for the simple effect of periods within each display indicated that in the integration task the response bias of observers in both configural groups remained stable over time, while the bias of observers in the non-configural display control group increased with time on task, F (2,23) = 6.5, p<.05. In the focused attention task, the only significant effect was for periods, such that observers became more conservative over time, F (2,108) = 9.5, p<.01.

An AVOVA on overall workload data indicated a significant effect for task type, F(1,90)=3.91, p<.05, such that the focused attention task (M=61.2) induced more overall workload than the integration task (M=53.4). There was also a main effect for display, F(2,90)=3.27, p<.05. Post-hoc comparisons revealed that the object configural display (M=51.6) induced significantly less workload than the non-configural display (M=63.9). The overall workload associated with the bar graph configural display (M=56.5) was not significantly different from that of the other two groups.

In addition to overall workload scores, ratings on the subscales of the TLX were also collected. These data were subjected to an ANOVA that revealed significant main effects for scales, F(4,385)=53.2, p<.001, display, F(2,90)=6.58, p<.01, and a significant 3-way interaction, task x display x scales, F(9,385)=3.20, p<.01. Separate task x display ANOVA's for each scale were computed to explore the 3-way interaction.

A significant display effect was observed for the Mental Demand scale, F(2,90)= 4.10, p<.05, and post-hoc tests revealed that the object configural display (M=48.0) induced significantly less mental demand than the non-configural control group (M=74.0) and the configural bar graph display (M=61.4). The two bar graph displays did not significantly differ from one another.

A significant effect on the Temporal Demand scale was observed for task, F(1,90)=6.82, p<.05, with the focused task (M=70.2) inducing higher temporal demand than the integration task (M=56.5).

For the Performance scale significant effects for display, F(2,90)=4.14, p<.05, and for the task x display interaction, F(2,90)=6.85, p<.01, were observed. Tests for the simple effect of display within each task revealed a significant effect for display in the integration task, F(2,45)=8.84, p<.01. Post-hoc tests indicated that the observers in the object display (M=22.2) or the bar graph configural display (M=28.4) groups did not differ significantly from one another, but that observers in both groups reported less performance workload than those in the control group (M=54.1). In the focused task there were no significant display differences (p>.05).

On the Effort scale the only significant effect was for display, F(2,90)=3.03, p=.05. Post-hoc tests indicated that observers in the object display condition (M=49.8) reported less effort than their cohorts in the non-configural display

group (M=65.8). Observers in the bar graph configural display condition (M=57.7) did not significantly differ from either group. No significant effects were observed for Frustration.

Participants' responses on the pre- and post-DSSQ were used to calculate scores on eleven scales using the formula: (post score – pre score)/SD, where the standard deviation used was obtained from a large normative group obtained from Matthews et al. (1999). A 2 (task) x 3 (display) x 11 (scales) ANOVA revealed a significant effect for scales, $F(6, 546)=33.05$, $p<.001$. All other effects failed to reach significance ($p>.05$ in each case). T-tests for each scale, using the Bonferroni correction, were computed to assess global pre-post vigil changes in stress state. This was done to verify that the absence of significant effects was not due to insensitivity of the DSSQ to the demands of the vigilance task. Scores on all the scales were significantly different from zero, with the exception of success motivation (M=-.19).

Thus, after the vigil participants reported reduced Energetic Arousal (M=-.64), Hedonic Tone (M=-.60), Intrinsic Motivation (M=-1.47), less able to Concentrate (M=-.65), and less Control and Confidence (M=-.84). They also reported feeling more tense (M=.45) and more Task-Related Cognitive Interference (M=.35) after the vigil. Observers also were less Self-Focused (M=-.42), had less Task-Irrelevant Cognitive Interference (M=-.53), and felt higher levels of Self-Esteem (M=.40) after the vigil. In terms of the secondary factors of the DSSQ, observers were less Task Engaged (reductions in Energetic Arousal, Intrinsic Motivation, and Concentration), and more Distressed (increases in Tense Arousal and reductions in Hedonic Tone). Note that the observed power for the predicted task x display x scales interaction was .7, which suggests that the lack of significance is not due to Type II error.

DISCUSSION

The results of the current study indicate that configural displays can attenuate the vigilance decrement, but only in an integration task. Hence, the performance effects of configural displays in vigilance appear to be task-dependent, consistent with the previous literature on these displays (Bennett & Flach, 1992). The facilitative effect of configural displays in integration tasks may result from the increases in signal saliency, which generally improves vigilance performance (Warm & Jerison, 1984). However, in the vigilance literature signal saliency is usually manipulated by varying the signal-to-noise ratio or the magnitude of the psychophysical discrimination (Warm & Jerison, 1984). The current results are not likely due to discrimination differences, since the physical discriminations were equivalent across displays. In this study, saliency is an emergent feature produced by the organization of display elements. As Buttigieg and Sanderson (1991) have noted, a clear definition of emergence has yet to be established, so linking it to quantifiable signal saliency effects is a difficult undertaking.

Display differences in response bias scores varied according to task. Within a task requiring the integration of information, differences among displays were observed as a function of time. Earlier in the watch participants in both configural display conditions responded more conservatively than those in the control condition, but by the last two periods, observers in the control group achieved levels of conservatism similar to those of the other groups. Configural displays may permit stable responding earlier in the vigil because they enhance situation awareness in integration tasks, allowing observers to more quickly adjust their responding according the frequency of signal appearance. Within the focused task there were no differences among display types, but in each condition conservatism increased over time. This likely reflects growing awareness on the part of observers that the frequency of critical signals is low (Craig, 1978).

Workload effects were influenced by task type as well as by the display employed, but these two factors did not interact. The object display reduced the overall workload associated with vigilance tasks regardless of whether the task required integration of information or focused attention. Hence, object configural displays may be employed to reduce the workload in both tasks, which is consistent with the argument that such displays can support performance on tasks requiring either integration of information or focused attention (Bennett & Flach, 1992).

The results for workload combined with the significant 3-way interaction between task, display, and periods for sensitivity indicate a complex pattern of associations and dissociations between workload and performance, depending on task and display type. Associations refer to cases in which increases in task load and increases in workload co-occur, while dissociations are cases in which increases in task load are linked with decreases in workload (or vice versa; see Parasuraman & Hancock, 2001; Yeh & Wickens, 1988).

Across all display formats the focused attention task induced more workload than the integration task, although sensitivity was statistically equivalent for the two tasks. This type of dissociation between workload and performance suggests that observers in the focused attention condition had to exert more effort in order to meet the performance of their cohorts in the integration task (Parasuraman & Hancock, 2001). Analysis of the subscales revealed that temporal demand may have been a major contributor to task differences. It may be that the requirement to scan three different elements in the display induced time pressure in observers experiencing the focused task, while those in the integration task may have been too preoccupied with the required mental calculations to attend to the temporal component of the task.

The performance-workload links across display formats were more complex and depended on task type. The object display reduced overall workload for both tasks but improved sensitivity only in the integration task. This indicates an association between workload and performance for that display in the integration task, but a dissociation between workload and performance in the focused task. The configural bar graph display showed a different pattern: Overall workload was not significantly reduced in either task, but sensitivity was enhanced in the integration task. This represents an association with respect to the focused task, and a dissociation with respect to the integration task (workload

was not significantly different from the control group but performance was improved).

The dissociation for the object display in the focused task (workload reduced but performance not significantly different from the control group) may result from an increase in resource investment in a data-limited task, which does not impact performance but increases perceived workload (Yeh and Wickens, 1988). In the configural bar graph integration task, however, the dissociation between overall workload and performance (workload was not affected but performance was enhanced) may result from similar working memory demands imposed by the two bar graph displays (Yeh & Wickens, 1988). The emergent feature of the bar graph configural display improves performance relative to a bar graph display without this feature, but it does not relieve the working memory demands associated with an integration task using bar graphs. In general, the workload results indicate that comparing object and bar graph configural displays only in terms of performance and concluding that they are functionally equivalent (Bennett & Flach, 1992; Buttigieg & Sanderson, 1991; Sanderson, et al., 1989) masks important differences between these two displays in how workload and performance are linked. In addition, it affirms the argument (Bennett & Flach, 1992; Buttigieg & Sanderson, 1991) that there is not necessarily a tradeoff between integrated and focused tasks with configural displays, as the object display reduced workload in both formats.

Analysis of subscale ratings on the TLX revealed patterns not observed in the evaluation of overall workload. Of particular interest is the task x display interactions observed for the performance scale. Display differences were observed for the integration task, with both configural displays reducing performance workload. This may be related to the enhanced situation awareness noted earlier in regard to observers' response bias scores.

The stress of sustained attention was not influenced by either task or display type, although overall pre-post changes on the scales of the DSSQ indicated that the vigil was stressful. These findings indicate that while configural displays eliminate the performance decrement and reduce workload (albeit in a complex way) they may not reduce the costs of that performance in terms of subjective stress. This is interesting, since previous studies have found that increasing signal salience decreases self-reports of stress in vigilance tasks. For instance, using the DSSQ, Temple et al. (2000) observed that participation in a vigilance task increased observers' stress levels, but only in the context of a difficult sensory discrimination. Apparently, in the current study, components other than the signal salience produced by emergent features contribute to the stress of monitoring in integration tasks as well as tasks requiring focused attention, and investigating those components is a matter for future research. Hancock (1998) and Scerbo (1998) have suggested that the context of vigilance tasks makes them aversive, since task conditions are imposed by either an experimenter or a manager in situations where the operator cannot easily quit. It may be that the absence of control by operators over their work tasks is a major contributor to the stress of vigilance. Regardless of the mechanism controlling the stress of sustained attention, the present study indicates that one cannot utilize configural displays for monitoring tasks with the expectation that operators will find the tasks less stressful.

REFERENCES

Bennett, K.B., & Flach, J.M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. Human Factors, 34, 513-533.

Buttigieg, M.A., & Sanderson, P.M. (1991) Emergent features in visual display design for two types of failure detection tasks. Human Factors, 33, 631-651.

Craig, A. (1978). Is the vigilance decrement simply a response toward probability matching? Human Factors, 20, 441-446.

Hancock, P.A. (1998). The price of freedom. Proceedings of the Human Factors and Ergonomics Society, 42, 1577-1578.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139-183). Amsterdam: North-Holland.

Matthews, G., Campbell, S.E., & Falconer, S. (2001). Assessment of motivational states in performance environments. Proceedings of the Human Factors and Ergonomics Society, 45, 906-910.

Matthews, G. Joyner, L., Gilliland, K., Campbell, S. Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), Personality psychology in Europe (vol 7), (pp. 335-350).

Maxwell, S.E., & Delaney, H.D. (1990). Designing experiments and analyzing data. Belmont, CA: Wadsworth.

Parasuraman, R., & Hancock, P.A. (2001). Adaptive control of mental workload. In P.A. Hancock & P.A. Desmond (Eds.), Stress, workload, and fatigue. (pp. 305-320), Mahwah, NJ: Erlbaum.

Sanderson, P.M., Flach, J.M., Buttigieg, M.A., & Casey, E.J. (1989). Object displays do not always support better integrated task performance. Human Factors, 31, 183-198.

Scerbo, M.W. (1998). Fifty years of vigilance research: Where are we now? Proceedings of the Human Factors and Ergonomics Society, 42, 1575-1576.

See, J.E., Howe, S.R., Warm, J.S., & Dember, W.N. (1995). A meta-analysis of the sensitivity decrement in vigilance. Psychological Bulletin, 117, 230-249.

Temple, J.G., Warm, J.S., Dember, W.N., Jones, K.S., LaGrange, C.M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. Human Factors, 42, 183-194.

Warm, J.S. (1993). Vigilance and target detection. In B.M. Huey, & C.D. Wickens, (Eds.), Workload transition: Implications for individual and team performance (pp.139-170). Washington, DC: National Academy Press.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 183-200). Hillsdale, NJ: Erlbaum.

Warm, J.S., & Jerison, H.J. (1984). The psychophysics of vigilance. In J.S. Warm (Ed.), Sustained attention in human performance (pp. 15-59). Chichester, UK: Wiley.

Yeh Y., & Wickens, C.D. (1988). Dissociation of performance and subjective measures of workload. Human Factors, 30, 111-120.

ACKNOWLEDGEMENT

The views expressed in this work are those of the author and do not necessarily reflect official Army policy. This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. The assistance of the Psychology Department at SUNY Farmingdale is also gratefully acknowledged.

Individual Differences in the Stress and Workload of Sustained Attention

James L. Szalma
University of Central Florida
Orlando, FL.

The effects of individual differences in dispositional pessimism and optimism, and choice of coping strategy, on performance, stress, and workload in vigilance tasks were investigated. Prior research indicated that pessimistic observers performed more poorly and experienced higher levels of stress than optimists. In addition, coping strategies employed by observers have been linked to the stress and workload associated with a variety of tasks. To date, no one has examined the relations among these variables in regard to vigilance within one study. Pessimism and optimism were found to be unrelated to performance, but predictive of both pre- and post-task stress and choice of coping strategy. There was also evidence that the influence of personality on post-task stress and choice of coping strategy may be mediated by pre-task state. Personality and coping strategies also influenced perceived workload.

Individual differences in operator characteristics influence the performance (Berch & Kanter, 1984) and stress (Helton, Dember, Warm & Matthews, 1999) of vigilance tasks. Moreover, operator characteristics such as personality traits and response strategy have been shown to influence perceived workload on a variety of tasks (Damos, 1988). A goal for the present investigation was to explore whether individual differences in observer coping strategies and in dispositional pessimism and optimism impact the performance, stress, and workload of monitoring tasks.

Matthews and Campbell (1998) noted that the study of coping strategies for dealing with workload and stress has received little attention in human factors. To address this gap they developed a coping inventory for task stressors (CITS) for post-task assessment of coping strategies individuals use while performing a task. Three strategies were identified based upon the transactional theory of stress (Lazarus & Folkman, 1984). These are task-focused coping, in which the individual copes with stress by trying to change the stressful situation; emotion-focused coping, aimed at changing one's emotional response; and avoidance-coping, in which the individual diverts attention away from the task.

Matthews and Campbell (1998) observed consistent patterns of correlation among coping strategies and the post-task stress states of observers, as measured by the Dundee Stress State Questionnaire (DSSQ), a multi-dimensional instrument for assessing transient states associated with mood, arousal, and fatigue (Matthews et al., 1999). The DSSQ consists of eleven factor analytically determined scales, including two recently established motivation scales (Matthews, Campbell, & Falconer, 2001), which load onto three secondary factors that reflect the individual's stress state. Three of the scales reflect Task-Engagement (Energetic Arousal, Concentration, and two Motivation scales (Intrinsic and Success); three scales constitute the dimension of Distress (Tense Arousal, Hedonic Tone, and Confidence and Control); and four scales define the dimension of Worry (Self-Focused Attention, Self-Esteem, Task Relevant Cognitive Interference, and Task-Irrelevant Cognitive Interference). Matthews and Campbell (1998) observed that task-coping correlated with Energetic Arousal and Motivation, indicating greater Task-Engagement, while emotion-focused and avoidance-coping were both related to increased Worry, reflected by both kinds of Cognitive Interference. Emotion-focused coping was also linked to higher Tense Arousal and lower Hedonic Tone (increases in Distress), while avoidance-coping was linked to a drop in Energetic Arousal (lower Task-Engagement).

They also reported positive correlations among coping strategies and perceived workload, as measured by the NASA Task Load Index (TLX; Hart & Staveland, 1988). Avoidance related to poor Performance ratings and lack of Effort, emotion-focused coping to Mental Demand and Frustration, and task-focused coping to increased Effort. Overall workload correlated most strongly with emotion-focused coping.

An additional goal for this investigation was to determine whether strategy choice is related to pessimism and optimism. Individuals high in pessimism tend to engage in emotion-focused coping, while optimists tend to use problem or task focused coping (Scheier & Carver, 1987). An experiment by Helton, Dember, Warm, & Matthews (1999) indicated that the performance and stress of sustained attention is moderated by pessimism and optimism. They reported that pessimists perform more poorly and find vigilance more stressful than optimists. Moreover, while Helton and his colleagues (1999) did not directly measure coping strategy, their findings that pessimists reported higher levels of Distress and less Task Engagement suggest that pessimists were employing emotion focused coping, since post-task Distress implies that form of coping, and optimists may have used task-focused coping, since higher Task Engagement implies use of that strategy (Matthews & Campbell, 1998).

METHOD

The present study was part of another investigation to compare the performance, stress, and workload associated with two kinds of configural displays, which contained emergent features to facilitate performance, relative to a

display that did not contain such features. These comparisons were made in the context of two task-types: an information integration task, in which observers were required to mentally combine information in the display to detect critical signals, and a focused attention task, which required observers to direct their attention to specific elements in the display. The performance, stress, and workload data obtained were analyzed nomothetically, without consideration of individual differences, and these results, with a detailed description of the displays, are described elsewhere in these proceedings (Szalma, in press).

Forty-eight male and forty-eight female undergraduates at SUNY Farmingdale participated in the study in exchange for extra credit. Upon entering the laboratory, participants completed the Optimism/Pessimism Inventory, which yields separate scores for these partially independent traits (Dember, Martin, Hummer, Howe, & Melton, 1989). After receiving instructions regarding the task participants completed the pre-task version of the DSSQ and then engaged in a 12-min practice vigil. The practice session was followed immediately by a 24-min vigil divided into four continuous 6-min periods of watch. After the vigil, participants completed a computerized version of the TLX, a post-task version of the DSSQ, and the CITS questionnaire. Administration of these measures was counterbalanced to control for potential order effects.

RESULTS

Preliminary analyses indicated that the individual difference variables in this study did not significantly interact with task or display type in influencing the performance or stress of the vigilance tasks. For workload, there were no interactions of individual difference variables with display type, and none of the higher order interactions (e.g. pessimism x task x display) were significant. However, there was an interaction of pessimism with task for overall workload, $F(1,84)=9.01$, $p<.01$, and the Performance subscale, $F(1,84)=3.98$, $p<.05$. Therefore the data were collapsed across task and display conditions for further analyses with performance and stress, and regressions of workload on the individual difference variables were computed separately for the two tasks.

Performance, in terms of signal detection theory measures of sensitivity (A') and response bias (B_d"), was calculated in two ways: 1) scores of observers were averaged across the four periods of watch for a measure of overall sensitivity and bias, and 2) a difference score between periods one and four was calculated to indicate the change in performance from the beginning to the end of the vigil. None of the regressions involving performance were significant ($p>.05$ in each case), indicating that neither pessimism/optimism nor coping strategy had a significant impact on performance.

Pre-Task Stress State

Results of regression analyses of each pre-task stress state on pessimism and optimism are shown in Tables 1 and 2. Both pessimism and optimism significantly predicted pre-task Distress and Task Engagement, but not pre-task Worry. Those high in pessimism tended to feel more Distress and less Engagement prior to the vigil, while those high in optimism reported less Distress and more Engagement.

Table 1. Summary of Regression Analysis for Variables Predicting Pre-task Distress

Variable	B	SE_b	β
Pessimism	.06	.01	.42**
Optimism	-.07	.02	-.36**

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. for pessimism: $R^2=.18$ ($p<.01$); for optimism $R^2=.13$ ($p<.01$); ** $p<.01$.

Table 2. Summary of Regression Analysis for Variables Predicting Pre-task Engagement

Variable	B	SE_b	β
Pessimism	-.04	.01	-.30**
Optimism	.04	.02	.27**

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. for pessimism: $R^2=.09$ ($p<.01$); for optimism: $R^2=.07$ ($p<.01$). ** $p<.01$

Coping Strategy

Pessimism significantly predicted use of emotion-focused coping, while optimism marginally predicted use of task focused coping. Avoidant coping was not related significantly to either personality type (See Tables 3-5).

Pre-vigil states significantly predicted choice of coping strategy (See Tables 3-5). Individuals high in pre-task Engagement preferred task coping and were less likely to use avoidant coping, while those high in Distress preferred emotion focused coping. Pre-task Worry was associated with both task- and emotion-focused coping.

Table 3. Summary of Regression Analysis for Variables Predicting Task-focused Coping Strategy

Variable	B	SE_b	β
Pre-Distress	-.68	.57	-.13
Pre-Engagement	1.53	.67	.24*
Pre-Worry	1.42	.47	.30**
Pessimism	-.13	.08	-.18
Optimism	.22	.1	.22#

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. For stress scales: $R^2=.15$ ($p<.05$). for pessimism $R^2=.03$ ($p=.08$); for optimism $R^2=.05$ ($p<.05$).
$.05<p<.1$; *$p<.05$; ** $p<.01$

Post-Task Stress State

For Post-task stress two sets of regressions were calculated: one in which pre-state was entered first and one in which coping strategies, pessimism, and optimism were entered into the regression without controlling for pre-task state. This was done so that the regressions could be compared

to determine whether the influence of traits or strategies on post-task stress was mediated by pre-task state. Neither pessimism nor optimism significantly predicted post-task Worry, but pessimism significantly predicted increased post-task Distress (b=.04, SE_b=.02, β=.24, p<.05). Optimism predicted increased post-task Engagement (b=.04, SE_b=.02, β=.24, p<.05) and decreased post-task Distress (b=-.05, SE_b=.02, β=-.24, p<.05). These regressions, however, were not significant when the corresponding pre-task state was entered first into the regression equation.

Table 4. Summary of Regression Analysis for Variables Predicting Emotion-focused Coping Strategy

Variable	B	SE_b	β
Pre-Distress	2.34	.47	.47**
Pre-Engagement	.37	.55	.06
Pre-Worry	1.38	.39	.31**
Pessimism	.22	.07	.31**
Optimism	-.21	.1	-.22

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. For stress scales R^2=.38 (p<.01); for pessimism R^2=.1 (p<.01); for optimism R^2=.05 (p<.05).
.05<p<.1; *p<.05; ** p<.01

Table 5. Summary of Regression Analysis for Variables Predicting Avoidant Coping Strategy

Variable	B	SE_b	β
Pre-Distress	.18	.63	.03
Pre-Engagement	-1.58	.75	-.23*
Pre-Worry	.25	.53	.05
Pessimism	.09	.08	.22
Optimism	.02	.11	.02

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. For stress scales R^2=.07 (p=.09); for pessimism R^2=.01 (p>.05); for optimism R^2<.001 (p>.05). *p<.05

Strategy was also related to reports of post-vigil stress: task-focused coping significantly predicted higher post-task Engagement (b=.05, SE_b=.02, β=.29, p<.01) and lower post-task Distress (b=.-06, SE_b=.02, β=.-29, p<.01), accounting for a significant proportion of variance after the pre-task state entered the equation (Engagement: b=.65, SE_b=.10, β=.54, p<.01; Distress: b=.28, SE_b=.12, β=.25, p<.05). Task coping also predicted increases in post-task Worry, (b=.05 SE_b=.02 β=.24, p<.05) although that relation was not significant when the effects of pre-task Worry were accounted for in the regression.

Emotion focused coping was linked to higher post-vigil levels of Distress (b=.07 SE_b=.02 β=.31 p<.01) and Worry (b=.05 SE_b=.02 β=.26 p<.01), and the relation to post-Worry remained after the variance due to the pre-vigil state was accounted for in the regression (Pre-Worry: b= .52 SE_b=.08 β=.54 p<.01), suggesting that emotion coping contributes to post-task Worry independently of pre-task Worry. However, the regression of post-task Distress on emotion-focused coping was non-significant when the effect of pre-vigil Distress entered the equation first (Pre-Distress: b=.28 SE_b=.12 β=.25 p<.05). Avoidance-coping significantly predicted loss of task-Engagement (b=-.04 SE_b=.02 β=-.21 p<.05), but these effects became non-significant when pre-task engagement was entered first into the equation (Pre-Engagement: b=.65, SE_b=.10, β=.54, p<.01). Avoidance was not significantly related to post-task Distress or Worry.

Workload

None of the coping strategies significantly impacted overall workload (p>.05 in each case). Pessimism significantly predicted increases in overall workload, but only in a focused task (see Table 6). In an integration task pessimism marginally predicted a reduction in workload. The prediction of decreases in workload with increases in optimism was also marginally significant in the focused task.

Table 6. Summary of Regression Analysis for Pessimism and Optimism Predicting Overall Workload in a Focused Task

	FOC			INT		
Variable	B	SE_b	β	B	SE_b	β
Pessimism	.89	.36	.34*	-.85	.48	-.25#
Optimism	-1.02	.55	-.26#	-.18	.59	-.05

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. Focused Task: R^2=.12 for pessimism (p<05); for optimism R^2=.07 (p>.05). Integration task: for pessimism R^2=.06 (p=.08); R^2= .002 for optimism (p>.05).
.1<p<.05; *p<.05; FOC=focused task; INT= integration task.

At the subscale level, only two scales showed significant regressions: Performance and Frustration. Task-focused coping predicted significantly less Performance workload in the focused task, and marginally predicted decreases in Performance workload in the integration task (see Table 7A). Neither of the other coping strategies was significantly related to Performance workload.

In a focused task, pessimism significantly predicted increased Performance workload in a focused task, and optimism marginally predicted Performance workload (see Table 7B). These traits were not significantly related to Performance in an integration task.

Table 7A. Summary of Regression Analysis for Coping Variables Predicting Performance Workload in Two Vigilance Tasks.

	FOC			INT		
Variable	B	SE_b	β	B	SE_b	β
Task coping	-1.57	.53	-.42**	-1.53	.85	-.27#
Emotion coping	-.11	.77	-.02	.96	.70	.20
Avoidant coping	-.18	.60	-.05	.11	.75	.02

Note. R^2=.17 (p<.05) for the focused task. R^2=.12 (p >.05) for the integration task. # .1>p>.05; ** p<.01; FOC=focused task; INT=integration task.

Table 7B. Summary of Regression Analysis for pessimism and optimism predicting Performance Workload in Two Vigilance Tasks

	FOC			INT		
Variable	B	SE_b	β	B	SE_b	β
Pessimism	.83	.40	.29*	-.22	.61	-.05
Optimism	-1.14	.60	-.27#	.68	.72	.14

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. Pessimism: R^2=.09 for the focused task, (p<.05); R^2=.003 for the integration task (p>.05); optimism: R^2=.07 for the focused task, (p=.06); R^2=.02 for the integration task (p>.05). # .1>p>.05; *p<.05 FOC=focused task; INT=integration task.

The only coping strategy related to Frustration was task-focused coping, which marginally predicted lower Frustration levels, but only in an integration task (see Table 8A). Pessimism significantly predicted increased Frustration, but only in a focused task (see Table 8B).

Table 8A. Summary of Regression Analysis for Coping Variables Predicting Frustration in Two Vigilance Tasks.

	FOC			INT		
Variable	B	SE_b	β	B	SE_b	β
Task coping	-.10	.63	-.02	-1.54	.92	-.24#
Emotion coping	1.42	.91	.23	1.22	.75	.23
Avoidant coping	1.12	.71	.24	.86	.81	.16

Note. R^2=.14 (p>.05) for the focused task. R^2=.19 (p<.05) for the integration task. # .1>p>.05; FOC=focused task; INT=integration task.

Table 8B. Summary of Regression Analysis for Pessimism and Optimism Predicting Frustration in Two Vigilance Tasks

	FOC			INT		
Variable	B	SE_b	β	B	SE_b	β
Pessimism	1.17	.45	.36*	.50	.67	.11
Optimism	-1.00	.71	-.21	-1.30	.78	-.24

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. Pessimism: for the focused task R^2=.13 (p<.05); for the integration task R^2=.01 (p>.05); optimism: for the focused task R^2=.05 (p>.05); for the integration task R^2=.02 (p=.10). *p<.05; FOC =focused task; INT=integration task.

DISCUSSION

The results of the above regressions indicates that coping strategy, pessimism, and optimism can influence the stress and workload of vigilance, and that pessimism and optimism can influence strategy choice. However, contrary to a previous investigation (Helton et al., 1999), observers achieved common performance levels regardless of their personality type or the coping strategy they chose. Thus, equivalent performance levels do not necessarily reflect similar mental states of operators, underscoring the importance of considering the effects of individual differences on stress and workload, as well as performance, in vigilance.

Both pre-task state and personality trait influenced operator choice of coping strategy. Individuals high in pre-task Engagement tended to prefer task-focused coping and tended not to use avoidant coping, while those who scored high on pre-task Distress tended to choose emotion focused coping. Observers with high pre task Worry reported that they employed both task and emotion-focused strategies. Individuals high in pessimism employed emotion focused strategies, while optimism marginally predicted use of task-focused strategy. The latter results are consistent with those of Helton et al. (1999), since pessimists in that study were more likely than optimists to feel higher Distress and lower Task Engagement. As noted earlier, Distress is associated with emotion focused coping, while Task Engagement is associated with task-focused coping (Matthews & Campbell, 1998).

Observers who engaged in task-focused coping reported less Distress and higher Task Engagement post-vigil, relations that persisted after the corresponding pre-task state was accounted for in the regression. This indicates that task-focused coping can reduce the stress of a vigilance task regardless of pre-task mood. Task-focused coping also predicted post-task Worry, but that relation became non-significant when pre-task Worry entered first in the regression. Recall that pre-task Worry predicted task-focused coping. It may be that individuals high in pre-task Worry were motivated to engage in task-focused coping as a way to reduce that state.

Emotion-focused coping predicted post-task Distress, but this effect became non-significant when pre-task Distress was accounted for in the regression. It is possible that individuals with high pre-vigil Distress were more motivated to use emotion-focused strategies in their attempts to deal with their negative affect.

Pessimism increased post-task Distress, while optimism decreased Distress and increased Engagement post-vigil. However, these effects were not significant when the pre-task states entered the regression first. This indicates that the influence of pessimism and optimism may be mediated by the cognitive state of the individual immediately prior to task performance. That is, pessimism and optimism might influence post-vigil stress state by influencing the operator's pre-task mood. Whether this mechanism generally applies to stress for tasks other than vigilance (or other types of sustained attention tasks) is a matter for future research.

None of the coping strategies significantly predicted overall workload, but increases in task-focused coping predicted decreased Performance workload. In addition, there were trends for increases in task-focused coping to decrease Performance workload and Frustration in an integration task. These results reinforce the argument that task-focused coping is more effective in ameliorating stress and workload effects associated with task performance compared to the other two strategies.

Pessimism influenced overall workload in the focused task, and marginally in the integration task, such that higher levels of pessimism and were associated with higher overall workload. There was also a trend for lower optimism to predict lower overall workload. The increase in overall workload for more pessimistic individuals may reflect fewer resources for pessimists due to interference by non-task

related processes, such as emotion focused coping to deal with their higher initial levels of Distress. Among the TLX subscales the only significant effects were for Performance and Frustration. In the focused task, pessimism was related to higher Performance workload ratings and increased Frustration, consistent with prior studies finding that pessimists tend to rate their own performance poorly and tend to report more negative affect (Scheier & Carver, 1987). The task difference may be due to the fact that all three focused tasks produced a performance decrement, while two of the integration tasks did not, and that the overall workload was higher in the focused attention task than in the integration task (Szalma, in press). That is, it may be that the effect of pessimism on overall workload emerges only in the context of tasks that consume more resources.

CONCLUSIONS

The results for the impact of coping strategy on stress and perceived workload were, in general, consistent with those of Matthews and Campbell (1998), indicating that the patterns of individual differences in coping they reported extend to vigilance tasks. The findings for coping and stress were also consistent with the findings of Helton et al. (1999), although the performance effects they observed were not found herein. Note, however, that they employed an extreme groups approach to categorize participants as pessimists or optimists, while in the present study these were treated as continuous variables. The latter approach has the advantage of avoiding the drawbacks associated with categorizing continuous variables (Pedhazur, 1997).

However, while personality appears to influence post-vigil stress and choice of strategy to cope with that stress, it may do so indirectly by influencing pre-task state. It may be that pessimists have fewer resources for dealing with the workload of vigilance, and this may be linked to their negative appraisals of their ability to successfully deal with the task. This is consistent with the finding that pessimists tend make more negative appraisals regarding their performance (Dember, et al. 1989; Scheier & Carver, 1987).

In sum, the relations described in this study also indicate that when designing or exploring methods of ameliorating the workload and stress of vigilance, both the coping strategies operators choose and their dispositional optimism and pessimism should be considered.

REFERENCES

Berch, D.B., & Kanter, D.R. (1984). Individual differences. In J.S. Warm (Ed.), Sustained attention in human performance (pp. 143-178). Chichester: Wiley.

Damos, D.L. (1988). Individual differences in subjective estimates of workload. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 231-237). Amsterdam: North-Holland.

Dember, W.N., Martin, S.H., Hummer, M.K., Howe, S.R., & Melton, R.S. (1989). The measurement of optimism and pessimism. Current Psychology: Research and Reviews, 8, 102-119.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139-183). Amsterdam: North-Holland.

Helton, W.S., Dember, W.N., Warm, J.S., & Matthews, G. (1999). Optimism, pessimism, and false failure feedback: Effects on vigilance performance. Current Psychology, 18, 311-325.

Lazarus, R.S., & Folkman, S. (1984). Stress, appraisal, and coping. New York: Springer-Verlag.

Matthews, G., & Campbell, S.E. (1998). Task-induced stress and individual differences in coping. Proceedings of the Human Factors and Ergonomics Society 42, 821-825.

Matthews, G., Campbell, S.E., & Falconer, S. (2001). Assessment of motivational states in performance environments. Proceedings of the Human Factors and Ergonomics Society, 45, 906-910.

Matthews, G. Joyner, L., Gilliland, K., Campbell, S. Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), Personality psychology in Europe (vol 7), (pp. 335-350).

Pedhazur, E.J. (1997). Multiple regression in behavioral research, 3rd ed. Ft. Worth: Harcourt Brace.

Scheier, M.F., & Carver, C.S. (1987). Dispositional optimism and physical well being: the influence of generalized outcome expectancies on health. Journal of Personality, 55, 170-210.

Szalma, J.L. (in press). Workload and stress of configural displays in vigilance tasks. Proceedings of the Human Factors and Ergonomics Society, 46.

ACKNOWLEDGEMENT

The views expressed in this work are those of the author and do not necessarily reflect official Army policy. This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. The assistance of the Psychology Department at SUNY Farmingdale is also gratefully acknowledged.

Individual Differences in the Stress and Workload of Sustained Attention

James L. Szalma

Abstract

The effects of individual differences in dispositional pessimism, optimism, and in choice of coping strategy on performance, stress, and workload in vigilance tasks were investigated. Prior research indicated that pessimistic observers performed more poorly and experienced higher levels of stress than optimists (Helton, Warm, Dember, & Matthews, 1999). In addition, coping strategies employed by observers have been linked to the stress and workload associated with a variety of tasks (Matthews & Campbell, 1998). To date, no one has examined the relations among these variables in regard to vigilance within one study. Pessimism and optimism were found to be unrelated to performance, but predictive of both pre- and post task stress state and choice of coping strategy. However, the influence of personality on post-task stress may be mediated by pre-task stress state. Personality and coping strategies also influenced perceived workload.

Introduction

The present study was part of another investigation to compare the performance, stress, and workload associated with two kinds of configural displays, which contained emergent features to facilitate performance, relative to a display that did not contain such features. These comparisons were made in the context of two task types: an information integration task, in which observers were required to mentally combine information in the display to detect critical signals, and a focused attention task, which required observers to attend to specific elements in the display (see Figure 1). The performance, stress, and workload data obtained were analyzed nomothetically, without consideration of individual differences, and these results are described elsewhere (Szalma, 2002).

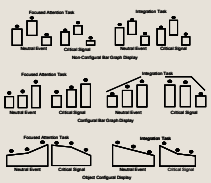


Figure 1. Displays employed in the present experiment. In each case the output is in the middle, flanked by the two inputs. The lines above the rectangles in the configural bar graph display indicate the emergent feature in the display. *Note*. The labels in each display and the lines above the rectangles are for illustrative purposes and were not present during the task.

Method

An equal number of male and female undergraduates (N=96) participated in the study in exchange for extra credit. Upon entering the laboratory, participants completed the Optimism/ Pessimism Inventory, which yields separate scores for these partially independent traits (Dember, Martin, Hummer, Howe, & Melton, 1989). After receiving instructions regarding the task participants completed the pre-task version of the Dundee Stress State Questionnaire (DSSQ; Matthews *et al* 1999; Matthews, Campbell, & Falconer, 2001) and then engaged in a 12-min practice vigil. The practice session was followed immediately by a 24-min vigil divided into four continuous 6-min periods of watch. After the vigil, participants completed a computerized version of the NASA-TLX (Hart & Staveland, 1988), a post-task version of the DSSQ, and the Coping Inventory for Task Stressors (CITS; Matthews & Campbell, 1998). Administration of the post-task measures was counterbalanced to control for potential order effects.

Results & Discussion

Individual difference variables did not significantly interact with task or display format in influencing performance or pre- or post-vigil stress (p>.05 in each case). There was an interaction between task type and pessimism for Overall Workload, $F(1,84)=9.01$, $p<.01$ and the Performance subscale of the TLX, $F(1,84)=3.98$, $p<.05$.

Therefore the data were collapsed across task and display conditions for further analyses with performance and stress, and regressions of workload on the individual difference variables were computed separately for the two tasks.

Performance

Performance, in terms of signal detection theory measures of sensitivity (A') and response bias (B_d'), was calculated in two ways: 1) Scores of observers were averaged across the four periods of watch for a measure of overall sensitivity and bias. 2) Difference scores between periods one and four were calculated to indicate the change in performance from the beginning to the end of the vigil. None of the regressions involving performance were significant (p>.05 in each case), indicating that the individual difference variables did not significantly influence performance. This is in contrast to results of Helton, Dember, Warm, & Matthews (1999), who observed that pessimists performed more poorly than optimists. Note, however, that Helton *et al* (1999) employed an extreme groups approach to categorize pessimists and optimists, while this study employed a regression approach.

Calculation of Stress-State & Coping Strategy Scores

Participant's responses to the pre- and post-DSSQ were used to calculate scores on eleven scales, using the formula: (score – scale mean)/SD, where the means and standard deviations were from a large normative group obtained from Matthews *et al*. (1999). The eleven scale scores were used to calculate three secondary factor scores: *Distress*, *Task Engagement*, and *Worry* (see Matthews *et al*. 1999).

Participant's responses to the CITS were used to calculate scores for three types of coping strategy: Task-focused coping, Emotion-focused coping, and Avoidant coping (see Matthews *et al*. 1999; see also Endler & Parker, 1990).

Regression Analyses

Regressions were calculated separately for optimism and pessimism because these traits are correlated (Dember *et al*. 1989). The correlation between pessimism and optimism in this study was r= -.47, p<.001.

For regressions of coping strategy on the pre-vigil DSSQ scales, the three secondary factors were entered into the regression equation in one step. Similarly, in regressions of workload and post-vigil stress state on coping strategies, the three coping scales were entered in one step.

Prediction of Pre-vigil Stress State from Personality Traits

Higher pessimism scores predicted increased pre-vigil *Distress* and decreased pre-vigil *Task Engagement*. Higher optimism scores predicted decreased pre-vigil *Distress* and increased pre-vigil *Task Engagement* (see Table 1).

Higher optimism scores predicted adoption of task-focused coping, while higher pessimism scores predicted adoption of emotion-focused coping (see Table 2).

Table 1. Summary of Regression Analyses for Variables Predicting Pre-vigil Distress and Pre-vigil Task Engagement

	Pre-vigil Distress				Pre-vigil Task Engagement			
	b	SE_b	β	R^2	b	SE_b	β	R^2
Pessimism	.06	.01	.42**	.18**	-.04	.01	-.30**	.09**
Optimism	-.07	.02	-.36**	.13**	.04	.02	.27**	.07**

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. ** p<.01.

Table 2. Summary of Regression Analyses for Pessimism and Optimism Predicting Coping Strategy

	Task-focused coping				Emotion-focused coping			
	b	SE_b	β	R^2	b	SE_b	β	R^2
Pessimism	-.13	.08	-.18	.03	.22	.07	.31**	.10**
Optimism	.22	.10	.23	.05*	-.21	.10	-.22	.05*

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. *p<.05; ** p<.01

Prediction of Coping Style from Pre-Vigil Stress State

➢High pre-vigil *Distress* predicted adoption of emotion-focused strategies. Individuals high in pre-vigil *Distress* may adopt that strategy to deal with their initial negative affect.

➢High pre-vigil *Task Engagement* and *Worry* predicted task-focused coping. Observers may be motivated to use task-focused coping as a way to reduce their worry.

➢Lower pre-vigil *Task Engagement* predicted higher levels of avoidant coping, suggesting that unmotivated observers are likely to adopt that strategy in the context of monitoring.

These findings (see Table 3) are consistent with those obtained by Matthews and Campbell (1998), although they did not observe a relation between task-focused coping and the *Worry* Dimension of the DSSQ.

Table 3. Summary of Regression Analyses for Pre-Vigil Stress State Predicting Coping Strategy

	Task-focused coping			Emotion focused coping			Avoidant coping		
	b	SE_b	β	b	SE_b	β	b	SE_b	β
Pre-Distress	-.68	.57	-.13	2.34	.47	.47**	.18	.63	.03
Pre-Engagement	1.53	.67	.24*	.37	.55	.06	-1.58	.75	-.23*
Pre-Worry	1.42	.47	.30**	1.38	.39	.31**	.25	.53	.05

Note. * p<.05; ** p<.01; For Task-focused coping: R^2= .15, p <.05.; For Emotion-focused coping: R^2= .38, p <.01.; For Avoidant coping: R^2=.07, p=.09.

Prediction of Post-Vigil Stress State

For Post-vigil stress two sets of regressions were calculated, one in which the corresponding pre-vigil state was entered first into the regression equation and one in which coping strategies and personality traits were entered into the regression without controlling for pre-vigil state. To simplify the analyses, each stress scale was analyzed ,min separate regressions.

Increases in pessimism predicted increased post-vigil *Distress*, and increases in optimism predicted lower post-vigil *Distress* and more *Task Engagement*, but these relations were not significant when the corresponding pre-task state is accounted for in the regression (p>.05 in each case).

➢Thus, personality influences post-vigil stress, but it may do so indirectly by influencing observers' pre-vigil state.

Task-focused coping was inversely related to post-vigil *Worry*, but this relation was not significant when pre-vigil *Worry* is accounted for in the regression (p>.05)

Higher emotion-focused coping scores predicted increased post-vigil *Distress*, but this relation was not significant when pre-vigil *Distress* was entered first into the regression (p>.05).

Task-focused coping predicted lower post-vigil *Distress* and higher post-vigil Task *Engagement*, after the corresponding pre-vigil state were accounted for in the regression. Emotion-focused coping predicted increased post-vigil *Worry* after pre-vigil *Worry* was entered first in the regression (see Table 4).

➢Thus, Task-focused coping may reduce the stress associated with vigilance. Emotion-focused coping increases at least one dimension of stress state.

ACKNOWLEDGEMENT

This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program . P.A. Hancock, principal investigator, administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the author and do not necessarily reflect official Army policy. The author wishes to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant. The assistance of the Psychology Department at SUNY Farmingdale is also gratefully acknowledged.

Table 4. Summary Regression Analyses for Coping Strategies Predicting Post-Vigil Stress State

	Post-vigil Distress					Post-vigil Task Engagement				
Coping strategy	b	SE_b	β	R^2	ΔR^2	b	SE_b	β	R^2	ΔR^2
Task-focused	.06	.02	-.30**	.5*	.09**	.05	.02	.29**	.51**	.10**

	Post-vigil Worry				
Coping strategy	b	SE_b	β	R^2	ΔR^2
Emotion-focused	.05	.02	.26**	.51**	.04**

Note. * p<.05; ** p<.01; R^2 is that obtained after both coping strategy and the corresponding pre-vigil stress state are entered into the regression. ΔR^2 reflects the increase in R^2 when coping strategy is entered second into the regression.

Prediction of Workload

Higher levels of pessimism predicted higher overall workload in a focused attention task, and marginally predicted higher workload in an integration task (see Table 5). Higher optimism marginally predicted decreased Overall Workload, but only in a focused attention task. At the subscale level, pessimism influenced Performance Workload and Frustration (see Table 6). This may be due to the tendency for pessimists to rate their own performance poorly, and report more negative affect (Scheier & Carver, 1987).

Task-focused coping reduced Performance Workload (see Table 6), consistent with findings of Matthews & Campbell (1998), although they found that such coping reduced other dimensions of workload as well. Whether task-focused coping is effective in reducing one or more dimensions of perceived workload may depend upon the nature of the task.

Table 5. Summary of Regression Analyses for Pessimism an Optimism Predicting Overall Workload in Two Vigilance Tasks.

	FOC				INT			
	b	SE_b	β	R^2	b	SE_b	β	R^2
Pessimism	.89	.36	.34*	.12*	-.85	.48	-.25#	.06#
Optimism	-1.02	.55	-.26#	.07	-.18	.59	-.05	.002

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. # .05<p<.1; *p<.05; FOC=focused task; INT= integration task.

Table 6. Summary of Regression Analyses for Pessimism, Optimism, and Task-Focused Coping Predicting Performance Workload and Frustration in a Focused Attention Task.

	Performance Workload				Frustration			
	b	SE_b	β	R^2	b	SE_b	β	R^2
Pessimism	.83	.40	.29*	.09*	1.17	.45	.36*	.13*
Optimism	-1.14	.60	-.27#	.07#	-1.00	.71	-.21	.05
Task-focused coping	-1.57	.53	-.42**	.17*	-.10	.63	-.02	.01

Note. Optimism and Pessimism were entered into separate regressions to avoid collinearity. # .05<p<.1; *p<.05; **p<.01; FOC=focused task; INT= integration task.

Task Differences

Task differences in regard to Workload might be due to: 1) the fact that all three focused attention tasks produced performance decrements, while only one of the integration tasks induced such a decline. 2) Overall Workload was higher in the focused attention task across all observers (see Szalma, 2002).

Conclusion

The coping strategies operators choose, and their degree of dispositional pessimism and optimism, influence the stress and perceived workload associated with vigilance tasks. These factors should be considered when designing systems requiring sustained attention.

References

Dember, W.N., Martin, S.H., Hummer, M.K., Howe, S.R., & Melton, R.S. (1989). The measurement of optimism and pessimism. *Current Psychology: Research and Reviews, 8*, 102-119.
Endler, N., & Parker, J. (1990). Multi dimensional assessment of coping: A critical review. *Journal of Personality and Social Psychology, 58*, 844-854.
Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.
Helton, W.S. Dember, W.N., Warm, J.S., & Matthews, G. (1999). Optimism, pessimism, and false failure feedback: Effects on vigilance performance. *Current Psychology, 18*, 311-325.
Matthews, G., & Campbell, S.E. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society 42*, 821-825.
Matthews, G., Campbell, S.E., & Falconer, S. (2001). Assessment of motivational states in performance environments. *Proceedings of the Human Factors and Ergonomics Society, 45*, 906-910.
Matthews, G., Joyner, L., Gilliland, K., Campbell, S. Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I.J. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe, 7*, 335-350.
Scheier, M.F., & Carver, C.S. (1987). Dispositional optimism and physical well-being: the influence of generalized outcome expectancies on health. *Journal of Personality, 55*, 170-210.
Szalma, J.L. (in press). Workload and stress of configural displays in vigilance tasks. *Proceedings of the Human Factors and Ergonomics Society, 46*.

The Impact of Stress on Adaptability in Teams

Heather A. Priest, C. Shawn Burke, Danielle Munim, & Eduardo Salas

MURI-OPUS

University of Central Florida

1 Abstract

Increased complexity has driven the need for researchers and practitioners alike to understand the effect of stress on team adaptability. Research has slowly begun to focus on the effect of stress on teams, what adaptability contributes to team performance under stress, and how to prepare teams to perform effectively when faced with stressors. Although these questions are just beginning to be investigated, several literature bases (e.g., team training, team effectiveness, individual adaptability) can be used to understand how teams can adaptively handle stress. Based on what we know about teams in general, individual adaptability, and how individuals perform under stress, we advance an initial framework to guide future research. This paper offers guidance pertaining to several theoretical issues that will be faced as we leverage past work in an attempt to understand team adaptability under stress. Furthermore, we also describe some practical training implications.

2 Introduction

•Environmental complexity and asymmetric warfare have caused an increased focus on the impact that stressors may have on team performance. Specifically, the question of how do these stressors differentially impact teams, as compared to individuals.

•An additional effect of complexity is that there is an increased need for teams to be adaptive if they are to be effective (Fowlkes, Burke, Wilson, Salas, & Pierce, in progress; Klein & Pierce, 2001).

•UCF is interested in both of the above areas. This is reflected within two research programs that are currently underway.The first program (funded by ARL) focuses on understanding team adaptability requirements. The second program (funded by ARO) focuses on the impact that stress has on adaptive and routine team performance.

3 Purpose

•To provide a preliminary theoretical framework that can used to guide our understanding of the impact of stress on team adaptability.

For more information, please contact Heather Priest at hpriest@ist.ucf.edu

4 Team Adaptability

Multiple Cues Push Adaptability

Environment (i.e., stressors) ⟷ Personnel (i.e., multicultural)

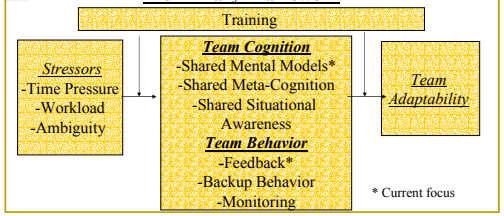
Adaptability: Requires the utilization of knowledge, skills, and attitudes that enable members to recognize deviations from expected action and then readjust actions accordingly.

5 What is Stress?

•Stress is often defined by the way it is operationalized.

Definition of Stress	Stressors
•"…any change in the environment that typically induces a high degree of unpleasant emotion (such as anxiety, guilt, or shame) and affects normal patterns of information processing" (Janis & Mann, 1977, p. 50). •An element of the environment that is detrimental to performance.	•Time Pressure •Workload •Ambiguity

6 Preliminary Framework



7 A Sampling of the Impact of Stress on Team Process & Adaptability

•In order for teams to perform adaptively under stress, certain team competencies must be trained and maintained. Although there are several competencies related to effective performance, due to time and space constraints, we choose to focus on shared mental models (SMM) and feedback.

Process	Propositions Regarding Stress	Contributions to Adaptability
•Shared Mental Models	Stress will effect the flexibility and updating of team SMMs and lead to a decrement in performance on event-based, teamwork tasks.	Enables teams to predict how other team members will behave.
•Feedback	Stress will impact the delivery, receipt, and acceptance of feedback.	Allows team members to adapt to novel outcomes and assist in designing new plans to better adapt

8 Sampling of Training: Preparing Teams To Adapt Under Stress

Strategy	What is Does?	Outcomes
Stress Exposure Training (SET)	• Presents trainees with knowledge of the stressors likely to be present and the impact on team behaviors •Provides strategies that members can use to mitigate the impact of stress •Provides practice in a relatively safe environment that gradually approximates the level of actual stressors.	•Reduced anxiety •Increased efficacy •Successfully applies skills to stressful environments •Improved cognition when stressors are present •Reduction of negative affect toward self and stressors

9 Future Considerations

•How do you create stressful environments within the lab? In the field?

•How can the impact of stress on team adaptability be studied using an event-based approach?

•What stressors have the biggest impact on team adaptability?

Applications of Fuzzy Signal Detection Theory to Responses to Terrorism

J.L. Szalma[1], P.A. Hancock[1], M. Mouloua[1], & R. Parasuraman[2]

1
MIT² INSTITUTE for SIMULATION & TRAINING
University of Central Florida
Multi-Disciplinary University Research Initiative
MURI-OPUS
Operator Performance Under Stress
www.mit.ucf.edu
2
The Catholic University Of America

Abstract

The detection of terrorists and terrorist activities in both military and civilian domains is central to the success of the war on terrorism. A recent advance in Signal Detection Theory (SDT) promises to enhance measurement of performance in such complex real world domains. This development, Fuzzy Signal Detection Theory (FSDT; Parasuraman, Masalonis, & Hancock, 2001), combines traditional SDT with Fuzzy Set Theory to extend signal detection analysis beyond the traditional crisp, categorical decision making model. FSDT allows for events to simultaneously be in more than one state category (e.g. signal and non-signal), so that the stimulus and response dimensions can be continuous rather than categorical. Consequently, FSDT can be employed in applied settings where the degree to which an event is a signal for detection may vary. This work aims to specify the ways in which FSDT can be applied to the problems faced by those designing systems to detect terrorist activity.

Fuzzy Logic

Traditional Set Theory: Set membership is mutually exclusive and exhaustive: elements either are or are not members of a set.

$$A \cap \overline{A} = 0$$

Fuzzy Set Theory: Events can simultaneously be and not be members of a set.

$$A \cap \overline{A} \neq 0$$

One assigns non-binary membership, or *degrees* of membership, to classes of events (fuzzification).

Signal Detection Theory

Assumes two possible states of the world: signal and noise (see Figure 1). This assumes a crisp, traditional set: events either are or are not signals.

State of the World	Response: Yes	Response: No
Signal	Hit (H)	Miss (M)
Noise	False Alarm (FA)	Correct Rejection (CR)

Figure 1. Relation between the state of the world and possible responses in traditional (crisp) SDT.

Signal Detection Theory Measures

Signal Detection theory permits the independent evaluation of perceptual sensitivity (e.g. d') and response bias (e.g. β; C). Sensitivity is expressed as the distance between the noise and signal plus noise distributions, while response bias is determined by where along the stimulus dimension (X in Figure 2a) the observer places the decision criterion. These indices can be visualized by constructing Receiver Operating Characteristics (ROC; see Figure 2b).

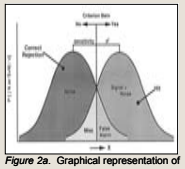


Figure 2a. Graphical representation of signal detection theory measures. Adapted from Swets (1996).

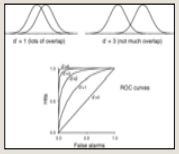


Figure 2b. Examples of Receiver Operating Characteristics. Adapted from Swets (1996).

Elements of Fuzzy Signal Detection Theory

Events can belong to the set "signal" (**s**) to a degree ranging from 0 to 1

Events can belong to the set "response" (**r**) to a degree ranging from 0 to 1

After defining these sets FSDT measures can be derived. (Parasuraman, Masalonis, & Hancock, 2000)

Four Steps for Computation of FSDT Measures

1) Select mapping functions for signal & response dimensions

2) Assignment of degrees of membership to the four outcomes (H, M, FA, CR) using mixed implication functions.

3) Compute fuzzy Hit, Miss, False Alarm, and Correct Rejection Rates

4) Compute detection theory measures of sensitivity and response bias

1. Mapping Functions

To assign degrees of (**s**, **r**) membership to events, all possible states of the world and each possible response must be evaluated using a mapping function. For the set **s** a mapping function relates the signal value (**s**) to a variable (or set of variables) that describe the state of the world (see Figure 3).

A mapping function for the set **r** relates the response value (**r**) to a response variable (or set of variables). For instance, a mapping function could be based upon confidence ratings of signal presence, a technique used in traditional SDT (Green & Swets, 1966;MacMillan & Creelman, 1991).

Note that the mapping function can be discrete or continuous, and can be derived empirically, theoretically, or based upon legal or industry standards. Either **s** or **r** (or both) can be fuzzy. The measurement challenge is to map **s** to the actual states-of-the-world. The derivation of the function may be theoretically based, but the reliability and validity of these functions should be established empirically.

2. Assignment of Fuzzy Set Membership to SDT Categories

Mixed Implication Functions

H = min (**s**,**r**)
M = max (**s**-**r**, 0)
FA = max (**r**-**s**, 0)
CR = min (1-**s**, 1-**r**)

- When **r>s**, some degree of FA membership will result, since the response is stronger than the actual degree of signal. Similarly, when **r<s**, some degree of miss membership results, since the response is less than the degree of "signalness" of the event.
- Performance is best when **r=s**, as the observer's responses are well mapped to the degree of signal present. This constitutes an ideal observer.
- Note that Parasuraman *et al.* (2000) did not claim these functions to be the only possible implication functions. The form of the mapping function can change with the domain of interest.

3. Computation of Fuzzy Hit and False Alarm Rate

$H = \Sigma(H_i)/\ \Sigma(s_i)$ for i=1 to N
$M = \Sigma(M_i)/\ \Sigma(s_i)$ for i =1 to N
$FA = \Sigma(FA_i)/\ \Sigma(1\text{-}s_i)$ for i=1to N
$CR = \Sigma(CR_i)/\ \Sigma(1\text{-}s_i)$ for i= 1 to N

Computational formulas obtained from Parasuraman *et al* (2000)

4. Computation of Fuzzy Sensitivity and Response Bias Measures

Fuzzy sensitivity and response bias measures can be calculated in the traditional way, since the fuzziness of signal and response have been captured in the computations of H and FA.
e.g. $d' = Z_H - Z_{FA}$; $c = -.5(Z_H + Z_{FA})$.

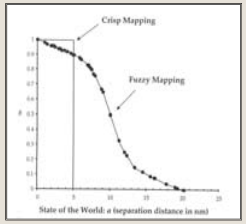


Figure 3. Sigmoid mapping function relating signal value, s, to each possible value of aircraft separation distance, a, as might occur in air traffic control. The rectangular function represents the standard or crisp SDT mapping of **s** to a. Adapted from Parasuraman, Masalonis, & Hancock (2000).

Applications of FSDT to Homeland Defense

There are several areas of application of FSDT to combating terrorism (see also Hancock & Hart, 2002).

Vigilance (e.g. airport luggage screening: what constitutes a weapon?)

Security Systems

Terrorist identification/detection
Preventing potential abuse: Ethnic Profiling. FSDT can be used to determine the degree to which response biases influence terrorist identification decisions.

Threat assessment (e.g. based upon intelligence data)

Combat (e.g. target acquisition; friend/foe identification)

Critical to these application will be identifying the mapping functions between the signal dimension and the state-of-the-world in a continuously changing environment in which the definition of a 'signal' changes over time and situations. In other words, the mapping functions may not be stable.

Continuous Signal Dimension with a Binary Response Dimension

In homeland defense applications, a common situation will be the case in which s is continuous but r is binary, since many situations related to security require specific actions, such as whether to open fire, detain a suspect (or a passenger), or sound an alarm.

In this case, the mixed implication functions reduce to:

If r=yes (1), then	If r= no (0), then
Hit = **s**	Hit=0
Miss = 0	Miss=**s**
FA = 1-**s**	FA=0
CR = 0	CR=1-**s**

Note that when **r** is binary, as **s** approaches maximal uncertainty (**s**=.5), FSDT could underestimate accuracy, but the fuzzification of **s** would improve estimation of response bias, since one could assess to what extent signal strength is related to response tendencies

Conclusions

The advent of Fuzzy Signal Detection Theory promises to improve the assessment of current and potential systems in response to terrorism. The domains in which this method can be applied reflect the spectrum of the war on terrorism, including security, analysis of intelligence data, identification of terrorists, as well as detection systems used in combat environments. FSDT accounts for the uncertainty not only in the observer, but in the state of the world itself. The days when events could be crisply categorized as signal or non-signal, if such times ever existed, are over. The grayness and uncertainty of threat level that permeates the post 9/11/01 world requires methods that capture that reality. FSDT offers a useful method to do so.

References

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42*, 636-659.
Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.
Hancock, P.A., & Hart, S. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in Design, 10*, 6-16.
Macmillan, N.A., & Creelman, C.D. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.
Swets, J.A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics.* Mahwah, NJ: Erlbaum

Acknowledgement

This research was supported by a Multidisciplinary University Research Initiative (MURI) program grant from the Army Research Office, P.A. Hancock, principal investigator. (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

A Human Factors Approach to the Investigation of Stress

H.C.N. Ganey, P.A. Hancock, J.L. Weaver, A. Greenwood-Ericksen, & M. Mouloua
University of Central Florida

Address For Correspondence: H.C.N. Ganey
Department of Psychology
University of Central Florida
Orlando, FL 32816

hganey@mail.ucf.edu

ABSTRACT

In this work, we discuss an integrated research program focusing on methods and strategies to support enhanced combat readiness and effectiveness in the face of myriad and numerous sources of operational stress. We base our approach on our existing general theory of performance under stress, enhancing this foundation (using the integration of a number of other theoretical structures) to provide quantitative models for stress prediction. Specifically, the model described below allows us to simultaneously capture the influences of both physical and cognitive forms of stress as well as their singular and interactive effects on response efficiency. This capability is critical for an area of psychological research dedicated almost solely to the study of inherently stressful occupations (e.g., law enforcement, aviation, military). Validation and refinement of these models is the focus of inter-related Laboratory and Field investigation, which will further focus on stress mitigation through support of the individual, support of the team, enhanced operator interfaces for operations under stress, and environmental manipulations for stress control.

Introduction

The study of stress has been and continues to be an important topic for psychological research. To date, the area has been investigated from a variety of theoretical perspectives and within different psychological sub fields (e.g., clinical, industrial-organizational). There has been some degree of sharing among the disciplines with regard to theoretical advances, and the benefit of this communication among sub fields is acknowledged. However, it should be noted that there are different emphases within disciplines due to the uniqueness of the situations studied within each. There are many theories and models concerning how stress affects human performance. Although some very useful and comprehensive models have been developed that provide explanatory power across these fields (e.g., Lazarus and Folkman's transactional model of stress) the nature of the situations studied within Human Factors demand a descriptive model that considers cognitive performance under physical demands. The long-accepted notion based on arousal was founded on an inverted-U description first promulgated by Yerkes and Dodson (1908). The popularity of this formulation is largely due to Hebb's (1955) influential Presidential address to the American Psychological Association. Unfortunately, the inverted-U description is lacking the specificity necessary to provide strong testable hypotheses. While its 'apparent' descriptive strength was partly responsible for its longevity as a putative explanation, in the end this weakness proved a too much to overcome and, therefore, became a significant factor in the demise of unitary arousal as a way to understand stress effects (Hancock 1987; Holland & Hancock, 1991; Hockey, 1983). Following upon the failure of unitary arousal, Hockey and his colleagues (see Hockey & Hamilton, 1983) provided a descriptive matrix that related various specific forms of environmental stress to sequential elements of human information processing. Strictly speaking, this taxonomic approach was descriptive rather than explanatory, but Hockey elaborated upon this notion in subsequent work to provide a full explanatory account (Hockey, 1997). Partly as a result of a highly influential meeting (i.e., Hockey, Gaillard, & Coles, 1986), a number of new theoretical formulations began to emerge. For example, Hancock (1986) provided early representations of the maximal adaptability model and Wickens (1986) tied stress to performance through the notion of gain on information-processing stages.

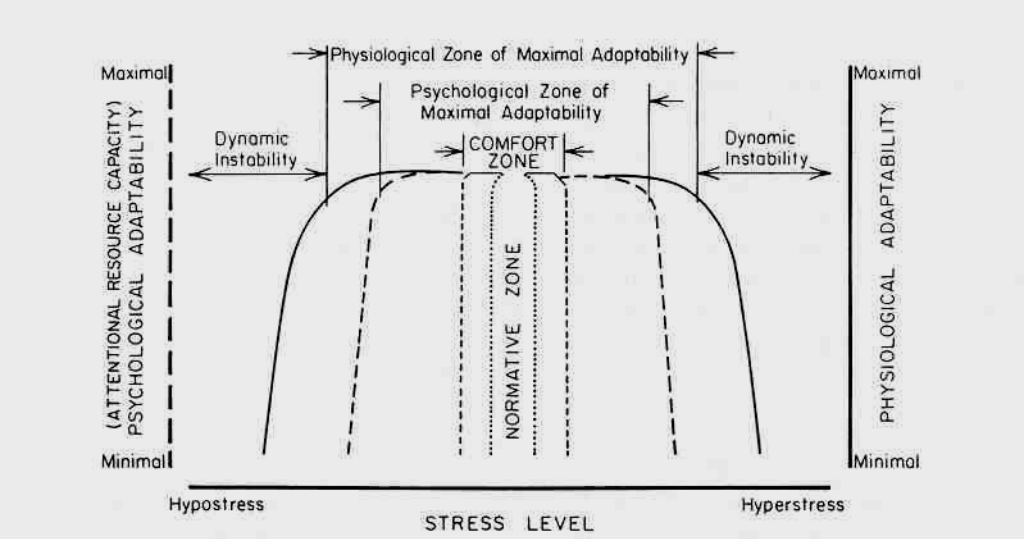


Figure 1. *The Maximal Adaptability Model of Stress and Performance***.**

The Maximal Adaptability Theory

The core description of the 'maximal adaptability' model upon which our program is founded is presented in Figure 1. As is evident from the vertical axes, this model provides a description of stress effects on both physiological and psychological performance. The general U-shaped form is preserved in this conception, except that it is better characterized as an 'extended-U.' Significant change in cognitive performance level is locked to uncompensated disturbance in physiological response. That is, cognitive performance is negatively affected when the individual is no longer able to maintain a homeostasis in respect to physiological effort (and see Weinstein, 1969). This provides a direct point of departure for research efforts on combined physical and mental workload. Failure in both physiological and psychological response can be thwarted by compensatory mechanisms for some period of time. Thus effort mobilization, in a general sense (where attentional 'narrowing' is a specific strategy in cognitive performance) permits prolonged tolerance of negative conditions. However, if neither exposure severity nor exposure duration are curtailed, the individual eventually experiences a failure state in which recovery becomes progressively more challenging (see Hancock, 1989, for an example of performance and workload under failure stress). Under such conditions, as attention is almost exclusively directed to the task, there can be drastic changes in the perception of time as attention to time-in-passing is exhausted. This is responsible for the common reports of time distortion under stress (see Hancock, 1997), which will be a major focus of our theoretical thrust. The maximal adaptability model also identifies the affective state of 'comfort' as an important baseline situation, and that aspect also will be considered as the antithetical situation to incipient failure. The maximal adaptability model therefore represents a framework from which to develop more specific predictive, mathematical models of stress effects.

Our Proposed Prototype Model

We have developed a prototype model of the effects of stress from the operational environment. The descriptive structure of this model is given in Figure 2. The first characterization is the input environmental demand of which task demand is a major component. As is evident in our model, task demand is addressed primarily by cognitive response, and overall 'environmental' demand by physiological response (although our model explicitly recognizes the cross-effects in this sequence). Physiological response to overall demand is divided into two components. First is a system specific component in which the proximally respondent system acts to meet the challenge posed by the environment and/or task. There are also concomitant effects that pervade the overall physiological response system. Successful resolution of demand by the proximal system results in automated response, maintenance of dynamic physiological stability and thus successful resolution. However, given that the demand exceeds the capability of the specific system, the response capability of the overall system is engaged. Successful general compensation leads to overall successful response through the same channel as specific system success.

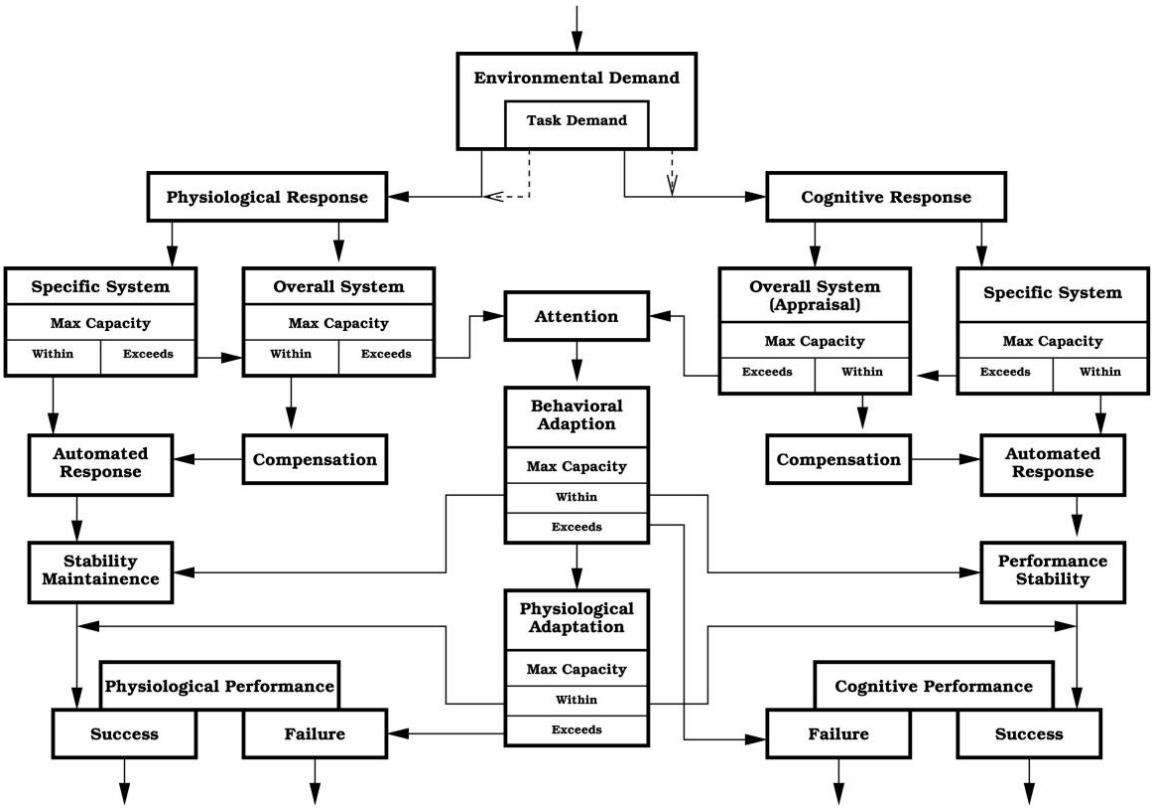


Figure 2. *A Prototype Process Model of Stress and Performance.*

Failure raises the question of physiological adaptation to the level of conscious attention and therefore requires some form of mitigational strategy, mediated by effortful decision-making. In fundamentally the same manner, the information processing system deals with task demands. First, they are addressed by proximal systems such as visual processing or auditory capacities. Failure at the specific level triggers a more encompassing appraisal of the situation, and in many ways this model incorporates both the unitary and distributed versions of the attentional resource models. Specifically, when collective resources are unable to process task demands using largely tactical, 'automated' responses, a higher level of attentional mediation is engaged, and a more strategic form of behavioral adaptation is adopted. Behavioral adaptation can range from increased conscious effort directed to particular aspects of the task, to such coping mechanisms as load shedding within a team situation. Load allocation can also be accomplished by changes in the task structure and interface configuration as mediated by the technological support available. When behavioral adaptations such as 'fight or flight' mechanisms are nullified by the circumstances, as is often the case in combat situation, the final form of response is within physiological adaptation. Although these have been illustrated as if they were lock-step relations, the reality of the situation is that they are highly parallel and highly interactive systems. We have, in the maximal adaptability model, provided a general indication of the nature of the failure function. Our first exploration has been to elaborate each of these representations along the path to the development of a fully articulated, mathematical representation of performance change.

The Measurement of Stress Effects

One of the first problems faced is the establishment of common measurement metrics by which to assess stress effects. We propose that there are three ways in which to solve such a problem. In the foundational model, Hancock indicated that stress can be viewed as an influence in three realms; the so-called 'trinity of stress.' These effects can be located as values in the environment itself, and are therefore physical measures, as is frequently the case in setting limits to occupational exposure (see Hancock & Vasmatzidis, 1999). The second locus of stress effects is in the interaction between such environments and

the specific individuals who are exposed to them. These represent the costs of adaptation, often referred to as 'strain effects. The final locus of effect, and perhaps most pragmatic way to measure stress effects, is directly through change in performance efficiency of the primary task to hand. Sometimes this is the easiest approach, at other times when future prediction is paramount, momentary changes in primary task performance are not so helpful. Each of these information sources in the 'trinity of stress' provides windows on the level of demand extracted by the environment. In the present work, we take the question of measurement one step further. We have engaged in a theoretical search for a more common currency to describe the level of demand. In our base model, these metrics are labeled information rate and information structure. They represent the spatial and temporal characteristics of the environment in relation to the capabilities of the exposed individual. As a relational measure, they are theoretically similar to the Gibsonian notion of an 'affordance.' Certainly, they will represent ratios, which will express the capability of the individual to respond versus the present level of demand. In our prototype model we can use such ratio measures as inputs to both predictive modeling and subsequent mitigation strategies such as adaptive systems (see Hancock & Chignell. 1995). The manner in which such measures can be used to assess the value of the model developed will be the central aspect of this first paper in a series reporting upon these pervasive 'energetic' effects.

Summary

The present paper introduces a program of research upon the effects of stress on operator performance. Specifically, the present work discusses the extant foundational model and the prototype that has been developed from this. We examine here, in detail, the fundamental problem of data capture and interpretation in inherently problematic collection conditions. In particular, we focus on how such measurements can address questions of concomitant physical and cognitive demand. Progress toward a solution of this tractable problem will permit refinement of the prototype into a fuller predictive model that can then be used for the examination of many different forms of 'energetic' effects on operator performance.

References

(All References Available on Request)

DOD Multidisciplinary Research Program: MURI Operator Performance Under Stress (OPUS).

White Paper

The Distortion of Perceptual Space-Time Under Stress

P.A. Hancock

J.L. Szalma

and

J.L. Weaver

Department of Psychology and the Institute for Simulation and Training

University of Central Florida

Work Performed under the U.S. Army Research Office: MURI: Optimizing Cognitive Readiness Under Combat Conditions

OPUS-MURI List of Collaborators

Universities:	Formal Connection: Kansas State University, University of Minnesota, Catholic University, University of Cincinnati. Informal Connection: Florida Institute of Technology, University of Toronto.
Industry:	United Defense, DCS Corporation.
Government:	NAWC-TSD
Individuals:	Dr. Jerry Krueger

THE DISTORTION OF PERCEPTUAL SPACE-TIME UNDER STRESS

Overview

We examine the distortion of perceptual space-time under stress and begin with the evaluation of the "narrowing" of attention that occurs in the presence of noxious or threatening conditions. This narrowing consists of a phenomenological reduction in the number and/or range of informational sources used by an individual as the severity of the stress increases. Early experimental investigations of this effect confounded sensory and attentional effects, largely because of an emphasis on the visual modality. More recent evidence indicates that 'narrowing' is an attentional effect and that cue salience can dictate the locus in the sensory field toward which such limited attentional resources are focused. Experimentation had overwhelmingly emphasized visuo-spatial capacities. Here we consider narrowing beyond the visual domain, with particular concern for distortion in time perception. Restricted attentional capacity, which results from stress, induces change in the strategic manner in which remaining resources are directed. Specifically, there is a reduction in environmental sampling, which results in a phenomenological distortion of perceived space-time. This adaptive response to dwindling resources remains effective in sustaining performance level only so long as such attention is directed to critical task-related cues. As new information is introduced into or is relocated in the environment, the adverse effects of resource limitations manifest themselves in reduced response efficiency. Theses effects are exacerbated by increasing the duration and/or the intensity of the stress. Evaluation of these propositions is on going. Confirmation of such hypotheses points to task-related and stress-related training as forms of mitigation as well as the dynamic reconfiguration of the task interface.

The Distortion of Perceptual Space-Time Under Stress

Time, like space, is a pure form of sensuous perception or intuition. It is the condition of all immediate active perception, as well as that which is perceived i.e. of all experience, and all which is experienced. Nature is made of space and time, and is a process. When we stress its spatial aspect, we are aware of its objective nature; when we stress its temporal aspect, we become aware of its subjective nature. As we perceive it, nature is an unending and continuous process of becoming. Things arrive and depart within time. Such things are not only within time, but they are also temporal—time is a way of existing.

Georg Wilhelm Friedrich Hegel
The Encylopedia of the Philosophical Sciences in Outline, 201

Introduction

Under extreme stress individuals report a 'narrowing' of the range of cues they extract from the surrounding environment. Reported first in anecdotes, this observation has received numerous experimental confirmations. Our purpose here is to explore this phenomenon, to examine the theoretical explanations of this effect and to frame this effect within a wider investigation of stress effects on human performance. We conclude with a series of problem statements that articulate experimental work needed to fully elucidate this phenomenon. We begin with existing experimental and theoretical information and show how each of these are only limited cases of the complete picture.

Distortion of Perceptual Space Under Stress

The first experimental report of 'narrowing' under stress was that of Bursill (1958) who examined the diminution of visual capacities in stressfully hot environments (and see Hancock & Vasmitzidis, 2000). However, the impetus for research in this area was provided not by these puzzling empirical observations but by the theoretical assertions of Easterbrook (1959), who indicated that emotional arousal restricts the utilization of the range of cues from the sensory environment. Easterbrook's proposal is spatial in conception implying that, under conditions of chronic or acute stress, peripheral stimuli are sacrificed at the expense of more centrally located cues. As attention narrows (i.e. as the number of cues attended to is reduced), performance capacity is preserved by the retention of focus on salient cues. Eventually, performance fails as stress increases and even salient cues become excluded. Given that extremes of under-load are as stressful as extremes of overload, it is evident that this proposition supports an inverted-U (Yerkes & Dodson, 1908), or an extended-U (Hancock & Warm, 1989) shaped function describing the relationship between stress level and performance efficiency. One assumption of Easterbrook's model is that there are irrelevant cues inherent in task performance, such that the task demands do not require the full resource capacity of the individual to perform. This problem comes from the lack of specificity in defining the lower limit to the number of cues used by an individual, as the precise meaning of the term, "fewer cues" remains undefined. In addition, performance may not retain stability if the process of eliminating task-irrelevant cues is improperly accomplished. Easterbrook's model does not describe the process that determines which cues are eliminated.

To support his model, Easterbrook (1959) discussed evidence that the perceived size of an important object is increased during states of emotional arousal. Intrinsic to this proposition, however, is the confound between the visual (sensory) contribution and the attentional (cognitive) contribution to narrowing, since Easterbrook did not indicate whether it is the center of the visual or the attentional field that holds primacy. Later work by Cornsweet (1969) produced evidence favoring an attentional explanation, showing that when peripheral cues were meaningful performance aids they were used by observers, even as arousal increased. Hancock and Dirkin (1983; see also Dirkin & Hancock, 1984) confirmed that 'narrowing' could occur to any part of the visual field, which implies that operators may be trained for cue selection under stress (Hancock & Dirkin, 1983).

In the late 60's, Wachtel (1967) argued that inconsistencies in empirical findings on the "narrowing" of attention resulted from a failure to distinguish between the two types of narrowing. He used a beam of light metaphor to express the concept. The width determines whether the observer focuses on a large or small area of perceptual space. Scanning represents the frequency with which a person scans the environment. Wachtel suggested that problems had arisen from the treatment of width and scanning aspects of attention as a unitary process. He also took issue with prior models of the inverted-U relationship, arguing that the downward arm resulted from a very narrow beam (unifocus), which scans a broad range of the environment (diffusion of attention). While agreeing with Easterbrook's explanation of narrowing in terms of cue utilization, he interpreted Easterbrook's model as a special case of a stationary beam where increases in arousal produce decreases in the width of that (stationary) beam. However, Easterbrook's hypothesis stipulated the negating of peripheral cues, which would therefore require "scanning" of the beam, and paradoxically Wachtel (1967) admitted that responsiveness to peripheral stimuli does indeed require scanning. Wachtel (1967) also suggested that the width and scanning aspects of attention can be partially controlled by an individual, depending on the person's ability to cope with his/her environment. Thus, narrowing occurs only when the individual perceives the event ("stressor") or threat as exceeding their resource capacity (see Lazarus & Folkman, 1984; Matthews, 2001).

Naatanen (1973) addressed the limitations of the inverted-U and hypothesized that sub-optimum performance is fostered by "divergent demands," rather than low or high activation levels. He proposed an increasing but negatively accelerating, function to describe the activation versus performance relationship and argued that the downward trend of the inverted-U resulted from a change in the appropriateness of the behavior relative to task goals. Inappropriate responses occurred through divergent demands that compete for and capture attention of the observer. These demands include secondary tasks as well as other external stimuli and internal states (e.g. anxiety and motivational factors).

There are problems with Naatanen's assertion. For instance, he proposes a negatively accelerated function for the activation-performance relation, but never supplied a rationale to support a negatively rather than positively accelerating function. Further, it's not clear that Naatanen considered time when claiming that performance improves "as long as the patterning of the response remains appropriate." His model does not address what would happen if a task lasted for 30 days, rather than the 30 second task he used as an example. Response may remain appropriate but whether performance would remain at criterion levels is undetermined.

"Narrowing" is a real world phenomenon that is not limited to laboratory conditions. Founded on both data and theory, the indications are that spatial perception is significantly influenced by stress. The question to be addressed next is whether time perception is also affected.

Distortions of Time Under Stress

We have concerned ourselves with the 'narrowing' of attention to spatial stimulation, largely confined to the visual modality. Here, we propose that:

i) There is 'narrowing' in the temporal dimension in the same way that there is 'narrowing' in the spatial dimension, and

ii) These are one fundamental process, being collectively distortions of spatio-temporal perception.

Like spatial 'narrowing,' we do not understand enough about time distortion to make it appear on demand. We do know that the content of an interval affects the perception of the length of that interval (cf., Block, 1990; Ornstein, 1969), and when social and environmental cues are minimized in sensory or perceptual deprivation, we can systematically influence the perception of duration (e.g., Banks & Cappon, 1962; Hancock, Arthur, Chrysler, & Lee, 1994). For stress-induced distortions in perceived time, there are several levels of evidence. These range from anecdotal accounts of 'private' experience during emergency conditions to empirical evidence from early explorations of the influence of danger on time perception.

Understandably, there are few examples of experiments that systematically evaluate the distortion of time under stress. One study on individuals with arachnophobia showed significant change in the appreciation of a specific time interval of exposure to a spider (Watts & Sharrock, 1984), while, Langer, Wapner, and Werner (1961, see also Langer, Werner, & Wapner, 1965; Werner & Wapner, 1955) showed that the threat of falling down a stairwell induced participants to alter their perception of a common time interval, relative to a no threat condition. Using a real-world stress of acceleration in a centrifuge (3g), Frankenhaeuser (1959; 1960) reported that participants produced underestimated intervals. She also observed that this difference grows proportionately with the length of the estimate. Such distortions in perceived time have also been observed in field studies. For example, Jurkovich, Campbell, Padtra, and Luterman (1987) found that under high stress, paramedics consistently underestimate the length of time spent at an emergency scene.

Theoretical Account of Time Distortion

Phenomenologically, individuals who experience 'time distortion' under stress report one of two forms of experience. For one group, everything appears as a 'blur' as though events were stacked together almost into one temporal 'moment' (see Fair, 1984). A second pattern has the individual reporting many different events with startling clarity as time itself appears to 'slow down.' Each of these patterns represent effects of 'time-in-memory' that are evident in the memory for the event. In these retrospective accounts, individuals are recalling events rather than referring to their experience in passing. Such immediate experience is referred to 'time-in-

passing' (see Roeckelein, 2000). In stress conditions sufficient to induce time distortion, the increase in event registration must be substantial. For such an individual, recalling the incident in memory is characterized by a recognition of the time distortion itself and a clarity of recall of many events which, since they cannot be reconciled with common time experience, are reported as being outside the normal run of behavior, e.g., "time seemed to slow down."

Let us consider the contrasting case. When attention is not directed toward the environmental but instead re-directed toward 'internal' events, few if any events are registered. Without sufficient events to fill an epoch, time seems suddenly 'speeded'. Thus, the phenomenological appearance of time co-varies with the rate of event registration. To the individual recalling such an incident time would appear to have speeded up and everything would have gone past 'in a blur'. These two patterns of behavior, both for time-in-passing and time-in-memory, accord well with the accounts reported by Fair (1984). The next step is to understand how stress produces these reactions, and for this we have to link stress reaction to attentional strategy. There is already a first link in the chain when we associate the direction of attention to external events as a form of 'fight' reaction, while withdrawal from external events may be considered a 'flight' reaction (and see Hancock & Weaver, 2002).

What can account for these striking observations of time distortion when under extreme stress or danger? We suggest that a causal account can be constructed through reference to our current understanding of how stress interacts with attention. The model proposed by Hancock and Warm (1989) provides this linkage by indicating how stress acts to reduce attentional capabilities, working in a similar manner to the way sources of environmental disturbance tax physiological response capacity. What Hancock and Warm (1989) did not explore was the process of failure as attentional capability waned. The first stage of failure is not failure at all but rather a more economic and efficient use of remaining attentional resources. A stress increases, however, attention narrows, performance suffers, and distortions between spatial and temporal perceptions occur.

Stress and the Distortion of Perceptual Space-Time: Avenues for Future Research

Above we have focused on distortion under stressful conditions, and following Hancock and Weaver (2002), we argued that these distortions have a common attentional mechanism. Two implications of this assertion are 1) that events (internal or external) that distort one dimension will distort the other, and 2) that these distortions are unlikely to be orthogonal. With very few exceptions, little research has addressed the possibility of an interaction between distortions of spatial and temporal perceptions in stressful situations on operator performance. Here, we discuss several avenues for future research aimed at addressing these issues, including the characteristics of the environment, the task, and the observer that influence these phenomena.

Interactive Effects. One initial question regarding the space-time distortions and their effects is the extent to which, under stressful conditions, perceptual distortions of space and time covary. One way to address this is to examine judgments of such stimuli that are made simultaneously vs. individually (in separate tasks), and to examine how these judgments change over time. In addition to comparing perceptual judgments, one can also evaluate the effects of stress on performance on information processing tasks that require one or both kinds of perceptual discrimination (e.g. target detection tasks). An important issue is whether we are dealing with

'time-in-memory' or time-in-passing (and to some extent space-in-memory vs. space-in-passing). Thus, the way in which perceptions of space and time interact to influence operator state will depend upon how temporal perceptions (and spatial perception, for that matter) are measured.

Mental Resources. If perceptions of space and time are related, they may draw on a common attentional resource pool, an issue we address in a subsequent white paper. The concern here is how the structure and function of resources relate to spatio-temporal perception. This can be evaluated through tasks in which spatial and temporal judgments are required but one task is secondary to the other. The demands of the primary task can then be manipulated to determine the effect on performance on the other task. If the tasks do draw on the same resources, increases in demand, e.g. a spatial task, would result in declines in a secondary task requiring a temporal discrimination. Individuals may also adopt resource allocation strategies to compensate for one distortion more than the other and the kinds of strategies they choose to cope with such interactive effects may be different from those they choose when coping with space or time distortion in single task conditions. The degree to which such differences can be observed will depend on whether the tasks employed are data or resource-limited (Norman & Bobrow, 1975), and the two types of discrimination may have different performance regions in which each task is data-limited.

A possible explanation for perceptual distortions under conditions of heavy workload and stress concerns the failure to switch tasks when appropriate. Some of the problems observed in secondary task methodology, such as ensuring that participants divide their time as instructed (e.g. 70% to the primary task and 30% to the secondary task) may result from the participant's inability to accurately judge how long they have attended to each task during a given time period. The degree to which distortions in perception of space-time are related to impairments in task switching under stressful conditions, and the degree to which these distortions are related to attention allocation strategies in a secondary task paradigm, are questions for empirical resolution.

Stress and Workload. The subjective experience of stress is multi-dimensional in character (Matthews, 2001; Matthews et al., 1999; Szalma, 1999). If space-time distortions share a common mechanism (attentional narrowing), then stress states (e.g. energy, tension, concentration, cognitive interference) associated with them should have similar profiles. Thus, a common "stress signature" should exist for the two perceptual dimensions. Similarly, their perceived workload profiles ought to be equivalent. Evidence for such hypotheses can be found in vigilance tasks, where both temporal and spatial tasks have similar workload profiles (Dittmar, Warm, Dember, & Ricks, 1993; Szalma et al., 1999). However, given that space-time distortions influence each other, the stress and workload profiles for combined space-time discrimination might be qualitatively different from that of a single discrimination condition.

Stressor Characteristics. Even if space and time do possess a common mechanism, it may be that specific stressors do not affect spatial and temporal perceptions in the same way. For example, heat and noise may distort both space and time, but not to the same degree or in the same fashion. It is important to note that spatial and temporal distortions may themselves be important stressors, since the tasks operators perform are the primary proximal sources of stress (Hancock & Warm, 1989). Consequently, some kinds of information processing might be more vulnerable to one or the other kind of perceptual distortion. Clearly, performance on tasks requiring spatial abilities, such as mental rotation, could suffer as a result of spatial distortion,

but they might be unaffected (or, in some cases, facilitated) by temporal distortion. Other tasks, such as those that rely heavily on working memory, mathematical ability, or tasks requiring target detection, could each show different patterns of change in response to space-time distortion.

Individual Differences. As with other aspects of perception, there are individual differences in the perception of space and time (Hancock & Weaver, 2002; Wachtel, 1967). Further, the subjective experience of stress is often multidimensional. Thus, it may be that two individuals are subjectively "stressed" by the same situation, but that their stress state profiles differ. Key processes likely to influence susceptibility to stress are motivational states, personality characteristics, cognitive style, and ability differences related to various information processes. Individuals are also likely to differ in the strategies they employ to cope with the distortions of space-time they experience while in a stressful environment, and these coping differences, if they exist, might depend on the quality (e.g. noise, heat, low S/N ratio) and source (e.g. environment, the task) of the stress and the personality traits of the individual.

Potential Benefits of Space-Time Distortion. Under certain conditions, the narrowing of spatial attention can benefit performance through the elimination of irrelevant cues. The precise conditions under which this occurs, however, remains unclear. In addition, it is important to identify the circumstances under which time distortion might actually prove beneficial. Here, operators perceive that they have additional time to complete the task at hand (Hancock & Weaver, 2002). This would have great benefit in task performance situations where attentional narrowing is less likely to have deleterious effects. At this point, this is an empirical question that might be amenable to controlled testing.

Summary

We have given only a brief overview of theory and research concerning distortional effects. Avenues for further research into the joint effects of space-time perception on human performance and mental state were presented and further research questions were gleaned. We specify these in more detail below.

Identified Research Questions

- Are perceptual distortions of space always accompanied by commensurate distortions of time?
- How do episodes of spatio-temporal distortions influence information processing capacity?
- Is attentional narrowing the common mechanism to account for spatial and temporal distortion?
- How is focusing on internal (body) vs. external (environment) events related to distortions of space and time?

- How does space-time perception relate to the range of resource consumption and allocation?
    - In a dual task situation, how do changes in the demand of a primary task affect performance on a secondary task?
    - How do spatial and temporal distortions influence task-switching behavior?

- Are the stress/workload profiles associated with distortions of perceived space and time similar?

- Do different environmental stressors influence distortions of space and time in the same fashion?

- What task factors are linked to space-time distortion?

- How are individual difference variables related to distortions in perceived space-time?

- What are the task and environmental circumstances in which space-time distortions facilitate performance?

Anticipated Outcomes

- Increased understanding of distortions of space-time will improve a general understanding of stress effects.

- A theoretical model of perceptual distortions can facilitate both experimental progress and modeling capacity.

- Identification of factors (task, environment, person) that control spatio-temporal distortions.

- Develop a profile of characteristics of individuals well suited for coping with stress and compensating for the debilitating effects of perceptual distortion. This will facilitate personnel selection.

- An exposition of the implications of both experimental and theoretical advances for the design of stress-adapted human-machine interfaces.

References

Banks, R., & Cappon, D. (1962). Effects of reduced sensory input on time perception. *Perceptual and Motor Skills*, **14**, 74.

Block, R.A. (1979). Time and consciousness. In: G. Underwood and R. Stevens (Eds.). *Aspects of Consciousness*. (pp. 179-217), Academic Press: London.

Block, R.A. (1990). Models of psychological time. In: R.A. Block. (Ed.). *Cognitive models of psychological time*. (pp. 1-35), Erlbaum: Hillsdale, NJ.

Bursill, A.E. (1958). The restriction of peripheral vision during exposure to hot and humid conditions. *Quarterly Journal of Experimental Psychology*, **10**, 113-129.

Caird, J.K., & Hancock, P.A. (1994). The perception of arrival time of on-coming vehicles at an intersection. *Ecological Psychology*, **6**, 83-109.

Carson, D.M. (1982). Temporal distortions and the ejection decision. *Flying Safety*, **38**, 8-28.

Carson, D.M. (1983). Temporal distortions and the ejection decision: An update. *Flying Safety*, **39**, 5-7.

Chapman, P.R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception*, **27**, 951-964.

Dember, W.N., Martin, S.H., Hummer, M.K., Howe, S.R., & Melton, R.S. (1989). The measurement of optimism and pessimism. *Current Psychology: Research and Reviews*, **8**, 102-119.

Dirkin, G.R., & Hancock, P.A. (1984). Attentional narrowing to the visual periphery under temporal and acoustic stress. *Aviation, Space and Environmental Medicine*, **55**, 457.

Dirkin, G.R., & Hancock, P.A. (1985). An attentional view of narrowing: The effect of noise and signal bias on discrimination in the peripheral visual field. In: I.D. Brown, R. Goldsmith, K. Coombes, & M.A. Sinclair (Eds.), *Ergonomics International 85: Proceedings of the Ninth Congress of the International Ergonomics Association,* 751- 753, Bournemouth, England, September.

Dittmar, M.L., Warm, J.S., Dember, W.N., & Ricks, D.F. (1993). Sex differences in vigilance performance and perceived workload. *The Journal of General Psychology*, **120(3)**, 309-322.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, **66**, 183-201.

Eysenck, H.J. (1959). Personality and the estimation of time. *Perceptual and Motor Skills*, **9**, 405-406.

Fair, D.G. (1984). *Temporal distortion*. Unpublished Manuscript, Department of Safety Science, University of Southern California. Los Angeles, CA.

Frankenhaeuser, M. (1959). *Estimation of time*. Stockholm: Almqvist and Wiksell.

Frankenhaeuser, M. (1960). Subjective time as affected by gravitational stress. *Scandinavian Journal of Psychology*, **1**, 1-6.

Gray, C.T., Gray, C.R., & Loehlin, J.C. (1975). Time perception: Effects of introversion/extraversion and task interest. *Perceptual and Motor Skills*, **41**, 703-708.

Hancock, P.A. (1993). Body temperature influences on duration estimation. J*ournal of General Psychology*, **120**, 197-216.

Hancock, P.A. (1997). *Essays on the future of human-machine systems*. Eden Prairie, MN: The BANTA Information Services Group.

Hancock, P.A. (2002). *Two-hundred stressful experiences in flight*. Unpublished Manuscript, University of Central Florida, Orlando, FL.

Hancock, P.A., Arthur, E.J., Chrysler, S.T., & Lee, J. (1994). The effects of sex, target duration, and illumination on the production of time intervals. *Acta Psychologica*, **86**, 57-67.

Hancock, P.A., & Dirkin, G.R. (1983). Stressor induced attentional narrowing: Implications for design and operation of person-machine systems. *Proceedings of the Human Factors Association of Canada*, **16**, 19-21.

Hancock, P.A., & Manser, M. (1997). Time-to-contact. In; A.M. Feyer (Ed.). *Occupational injury.* London: Taylor and Francis.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, **31**, 519-537.

Hancock, P.A., & Weaver, J.L. (2002, in press). On time distortion under stress. *Theoretical Issues in Ergonomic Science.*

Ikeda, M., & Takeuchi, T. (1975). Influence of foveal load on the functional visual field. *Perception & Psychophysics*, **18**, 255-260.

Jurkovich, G.J., Campbell, D., Padtra., & Luterman, A. (1987). Paramedic perception of elapsed field time. *Journal of Trauma*, **27**, 892-897.

Korchin, S. (1964). Anxiety and cognition. In C. Scheerer (Ed.), *Cognition: Theory, research, promise*, pp. 58-78. New York: Harper & Row

Langer, J., Wapner, S., & Werner, H. (1961). The effect of danger upon the perception of time. *American Journal of Psychology*, **74**, 94-97.

Langer, J., Werner, H., & Wapner, S. (1965). Apparent speed of walking under conditions of danger. *Journal of General Psychology*, **73**, 291-298.

Lazarus, R., & Folkman, S. (1984). Stress, appraisal, and coping. New York: Springer Verlag.

Matthews, G. (2001). Levels of transaction: A cognitive science framework for operator stress. In P.A. Hancock & P.A. Desmond (Eds.), *Stress, workload, and fatigue*. Mahwah, NJ: Erlbaum.

Matthews, G., & Campbell, S.E. (1998). Task-induced stress and individual differences in coping. In *Proceedings of the Human Factors and Ergonomics Society*, **42**, pp. 821-825. Santa Monica, CA: Human Factors and Ergonomics Society.

Matthews, G., Joynder, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state 'big three'? In I. Mervielde, I.J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe (vo. l 7),* pp. 335-350. Tilburg: Tilburg University Press.

Naatanen, R. (1973). The inverted-U relationship between activation and performance: A critical review. In: S. Kornblum (Ed.), *Attention and performance IV*, (pp. 155-174). New York: Academic Press.

Norem, J.K., & Illingworth, K.S.S. (1993). Strategy-dependent effects of reflecting on self and tasks: Some implications of optimism and defensive pessimism. *Journal of Personality and Social Psychology*, **65**(4), 822-835.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processing. *Journal of Cognitive Psychology*, **7**, 44-60.

Ornstein, R.E. (1969). *On the experience of time*. Penguin: London.

Roeckelein, J.E. (2000). *The concept of time in psychology: A resource book and annotated bibliography*. Greenwood Press: Westport, CT.

Szalma, J.L. (1999). *Sensory and temporal determinants of workload and stress in sustained attention*. Unpublished doctoral dissertation, University of Cincinnati, Cincinnati, OH

Szalma, J.L., Warm, J.S., Matthews, G., Dember, W.N., Weiler, E.M., Meier, A., & Eggemeier, F.T. (1999). Sensory and temporal determinants of workload and stress in sustained attention. In *Proceedings of the XIVth Triennial Congress of the*

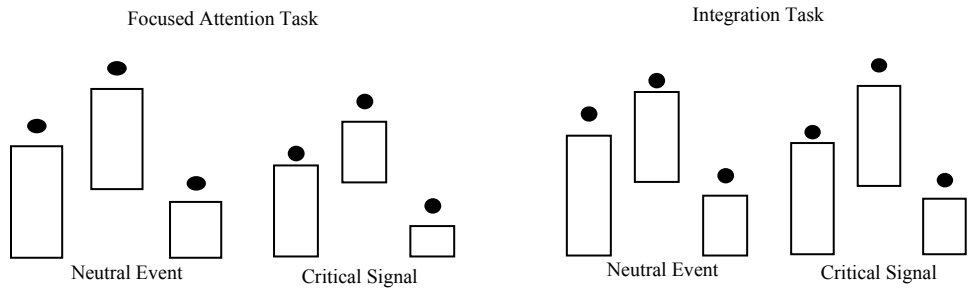
International Ergonomics Association and the Human Factors Society 44th Annual Meeting (pp. 3-386 – 3-389). Santa Monica, Ca: Human Factors and Ergonomics Society.

Wachtel, P.L. (1967). Conceptions of broad and narrow attention. *Psychological Bulletin*, **68**, 417-429.

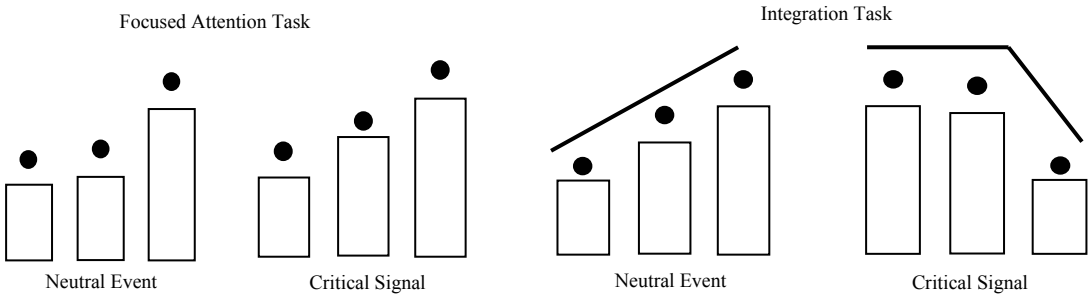
Watts, F.N., & Sharrock, R. (1984). Fear and time estimation. *Perceptual and Motor Skills*, **59**, 597-598.

Werner, H., & Wapner, S. (1955). Changes in psychological distance under conditions of danger. *Journal of Personality*, **24**, 153-167.

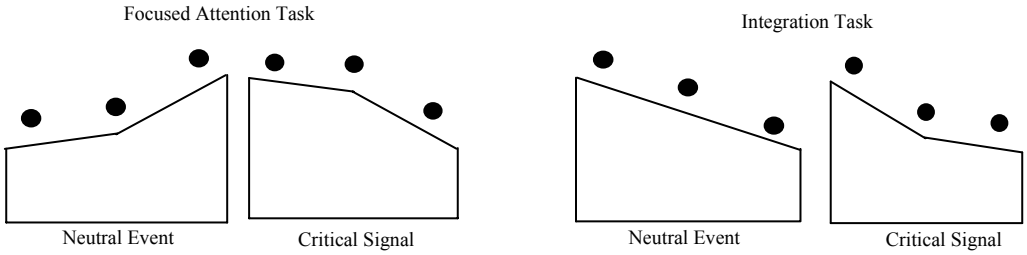
Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, **18**, 459-482.



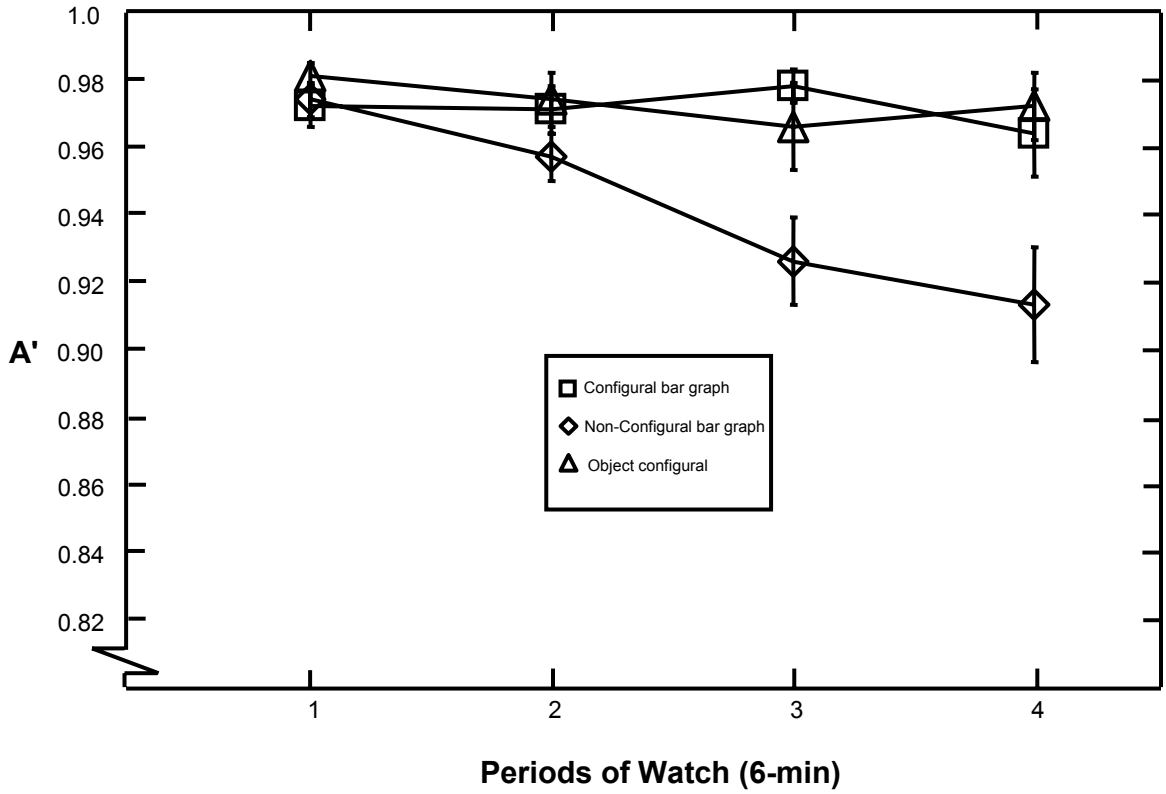
Non-Configural Bar Graph

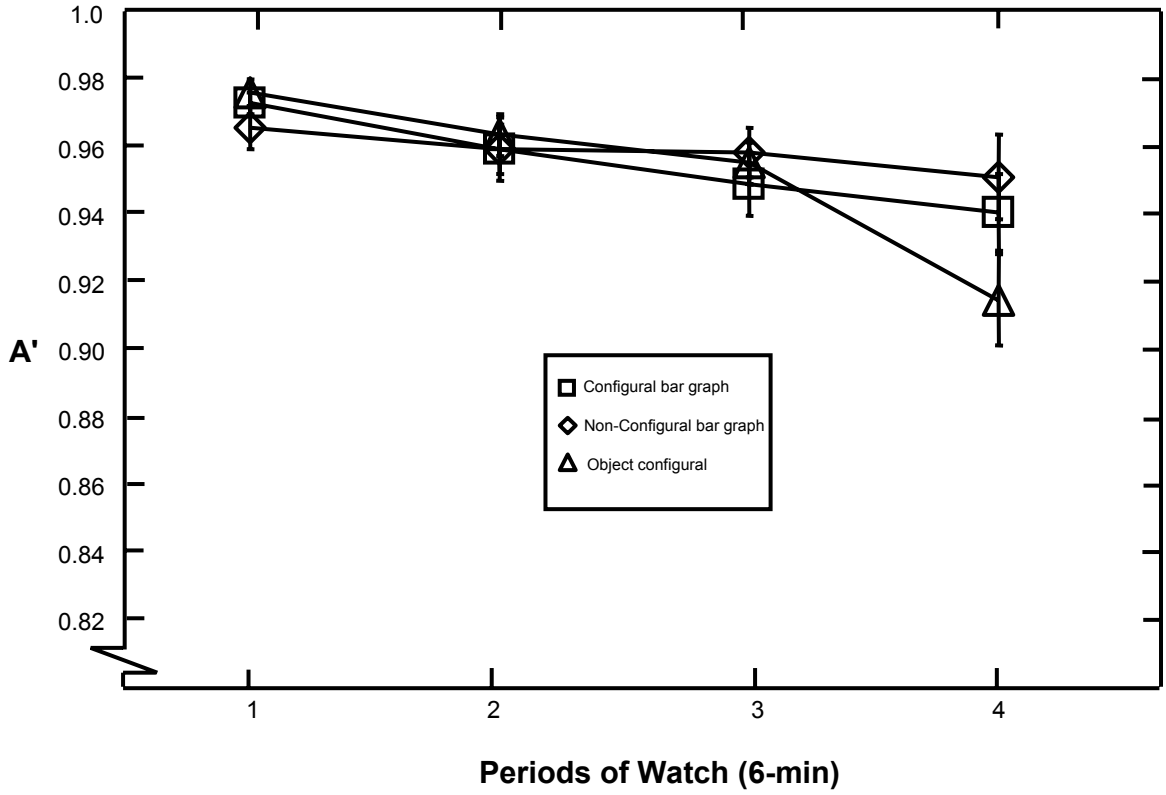


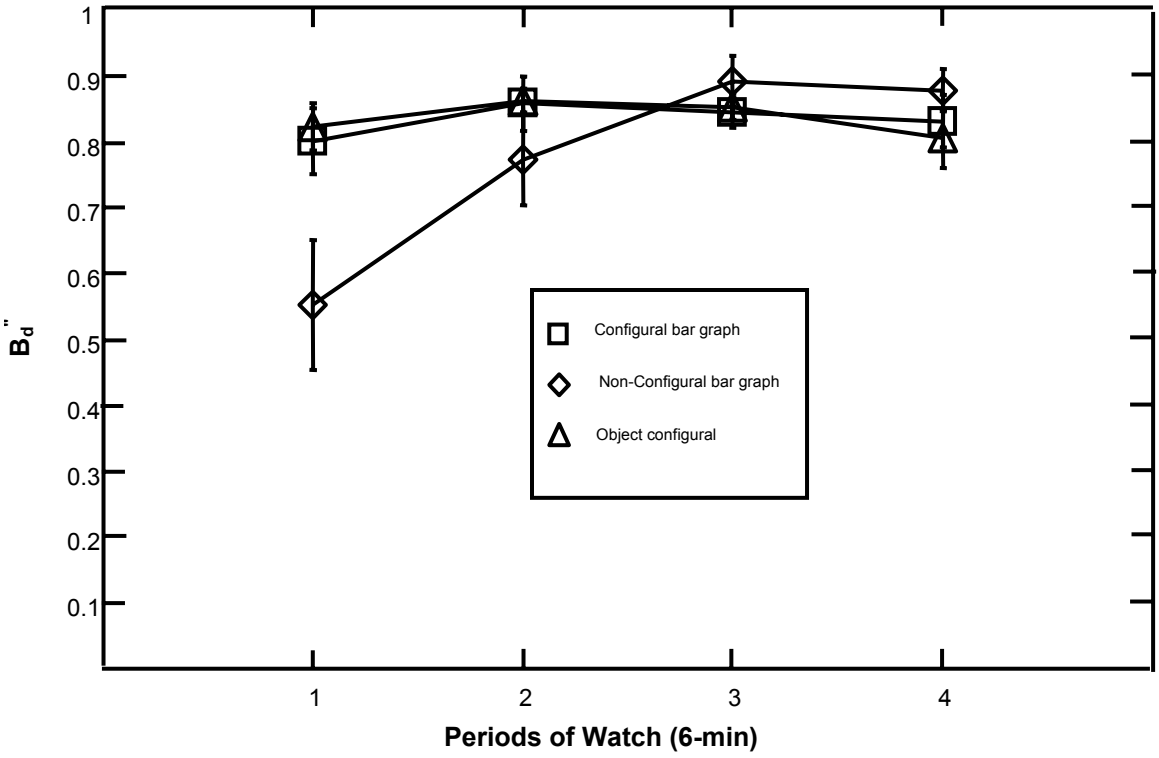
Configural Bar Graph

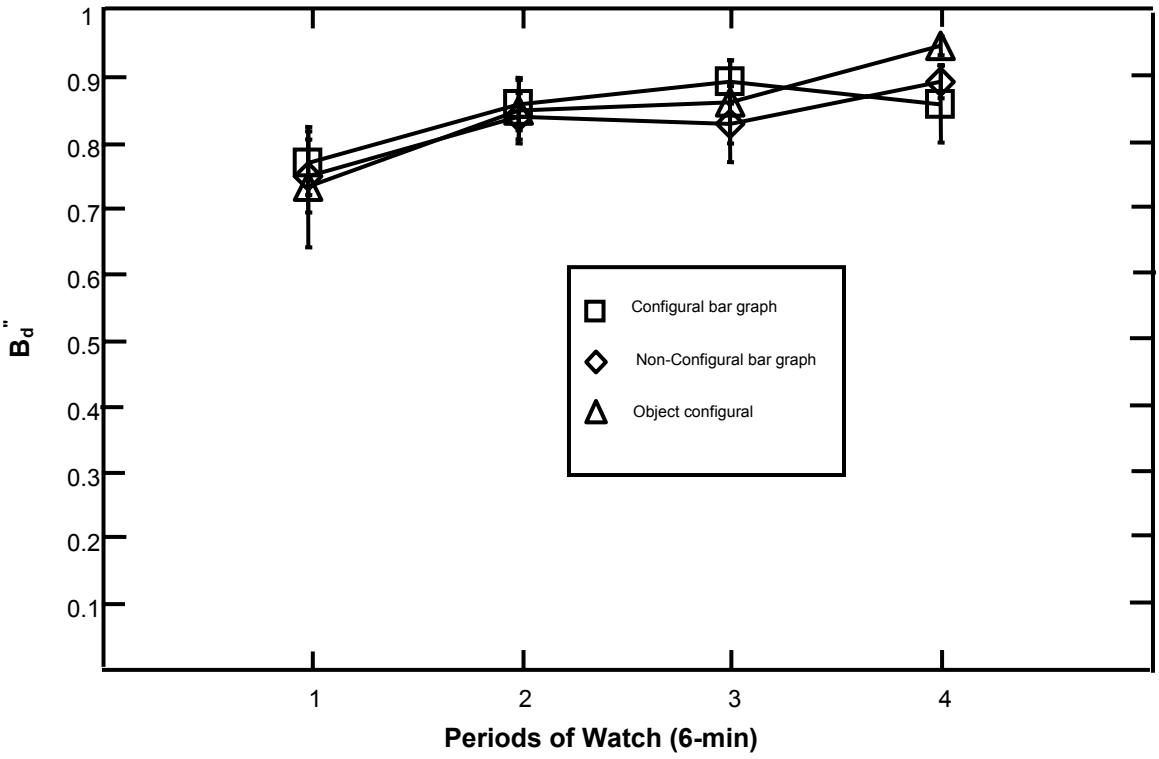


Object Configural


1.0
0.98
0.96
0.94
0.92
A'
0.90
0.88
0.86
0.84
0.82
1
2
3
4
Configural bar graph
Non-Configural bar graph
Object configural
Periods of Watch (6-min)


1.0
0.98
0.96
0.94
0.92
A'
0.90
0.88
0.86
0.84
0.82
1
2
3
4
Configural bar graph
Non-Configural bar graph
Object configural
Periods of Watch (6-min)


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
B_d''
Configural bar graph
Non-Configural bar graph
Object configural
1
2
3
4
Periods of Watch (6-min)


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
B_d''
Configural bar graph
Non-Configural bar graph
Object configural
1
2
3
4
Periods of Watch (6-min)

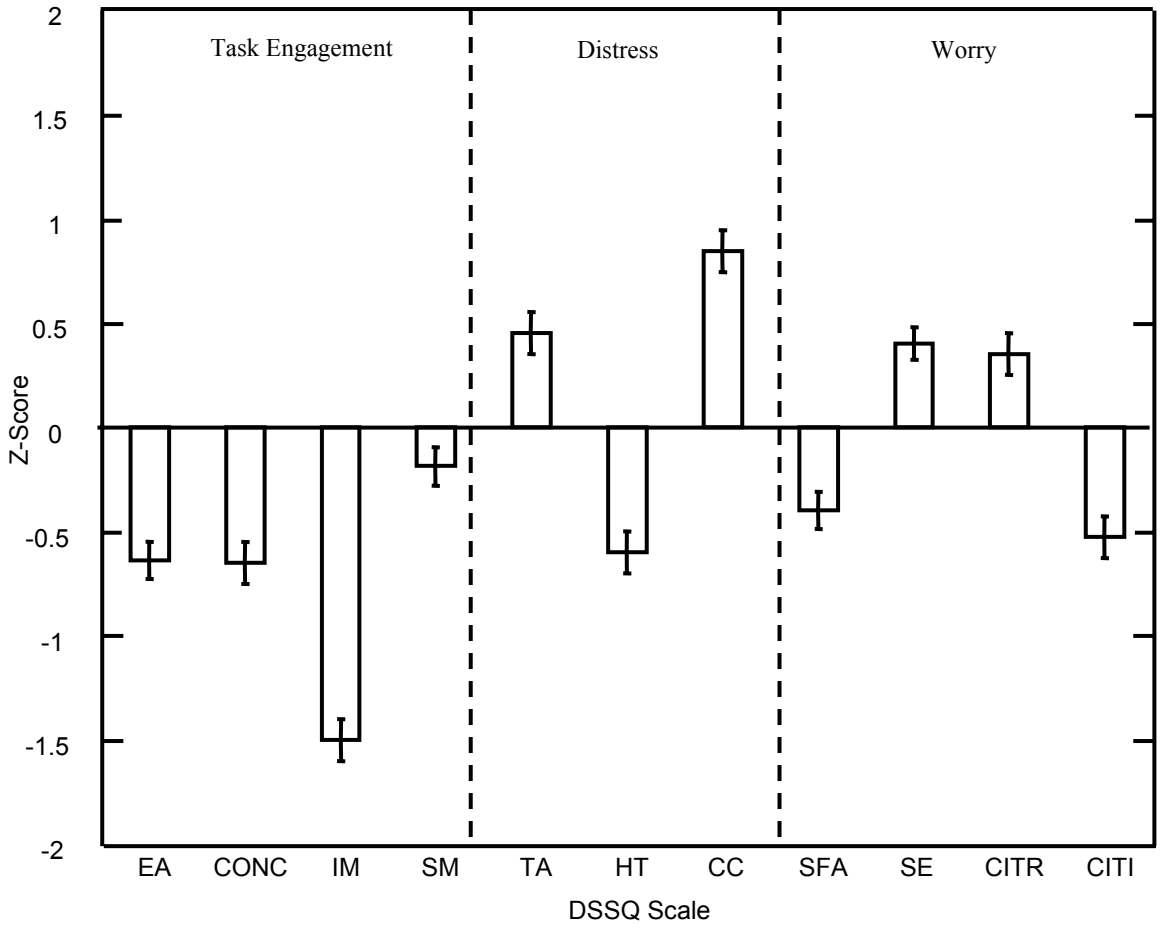

Task Engagement
Distress
Worry
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Z-Score
EA
CONC
IM
SM
TA
HT
CC
SFA
SE
CITR
CITI
DSSQ Scale

Table 1.

Means and Standard Errors for Three Coping Strategies for Each Experimental Condition.

	Task-Focused Coping		Emotion-Focused Coping		Avoidant Coping		
	Integration	Focused	Integration	Focused	Integration	Focused	*M*
Non-Con	15.6 (1.3)	16.8 (1.3)	7.6 (1.4)	6.8 (1.1)	7.6 (1.3)	5.0 (1.2)	9.9
Con-Bar	17.2 (1.1)	16.1 (1.5)	8.9 (1.2)	5.8 (.8)	7.9 (1.0)	6.1 (1.5)	10.3
Con-Obj	16.8 (1.1)	15.8 (1.3)	6.3 (1.5)	4.1 (.8)	7.6 (1.7)	5.1 (.9)	9.3
M	16.5	16.2	7.3	5.6	7.7	5.4	

Note. Standard Errors are in Parentheses. Scores on each scale can range from 0 to 28. Non-Con= Non-configural bar graph display; Con-Bar = configural bar graph display; Con-Obj= configural object display.

Team Performance Laboratory
University of Central Florida

Empirical Update:

We are currently collecting the final teams to conclude our study on the impact of time pressure on team process and performance. We have used the Flashpoint test-bed to examine this issue. This is a military based task most akin to a reconnaissance team task where sequential task interdependence is required, and the available cues in terms of team mates' actions are solely verbal. Concurrent with the wrapping up of this effort we are now pursuing a three-pronged approach to looking at the impact of stress on team process and performance. We have spent the past couple of months creating a set of three test beds that differ across some fundamental aspects of team type (reconnaissance, command and control, combat team), interdependence required (sequential, reciprocal, team), and type of cues that are received from teammates' (verbal only, verbal and low visual detail, detailed visual and verbal). The plan is to use these three test-beds to compare and contrast the findings related to not only how various stressors may impact team process and performance, but also to focus on the differential impact depending on team and task requirements. All three test-beds are based on military tasks. We are now beginning to collect data in two of the test-beds (DDD, GhostRecon), using the same manipulation of workload and measures within both. When the initial Flashpoint study concludes, this test-bed will also follow suit with the workload manipulation.



University of Central Florida

Combat Readiness and Stress: Laboratory Investigations of Teams

Clint Bowers, PhD., Eduardo Salas, PhD., Kia Asberg,
Shawn Burke, PhD., Heather Priest, MS., and Laura Milham.

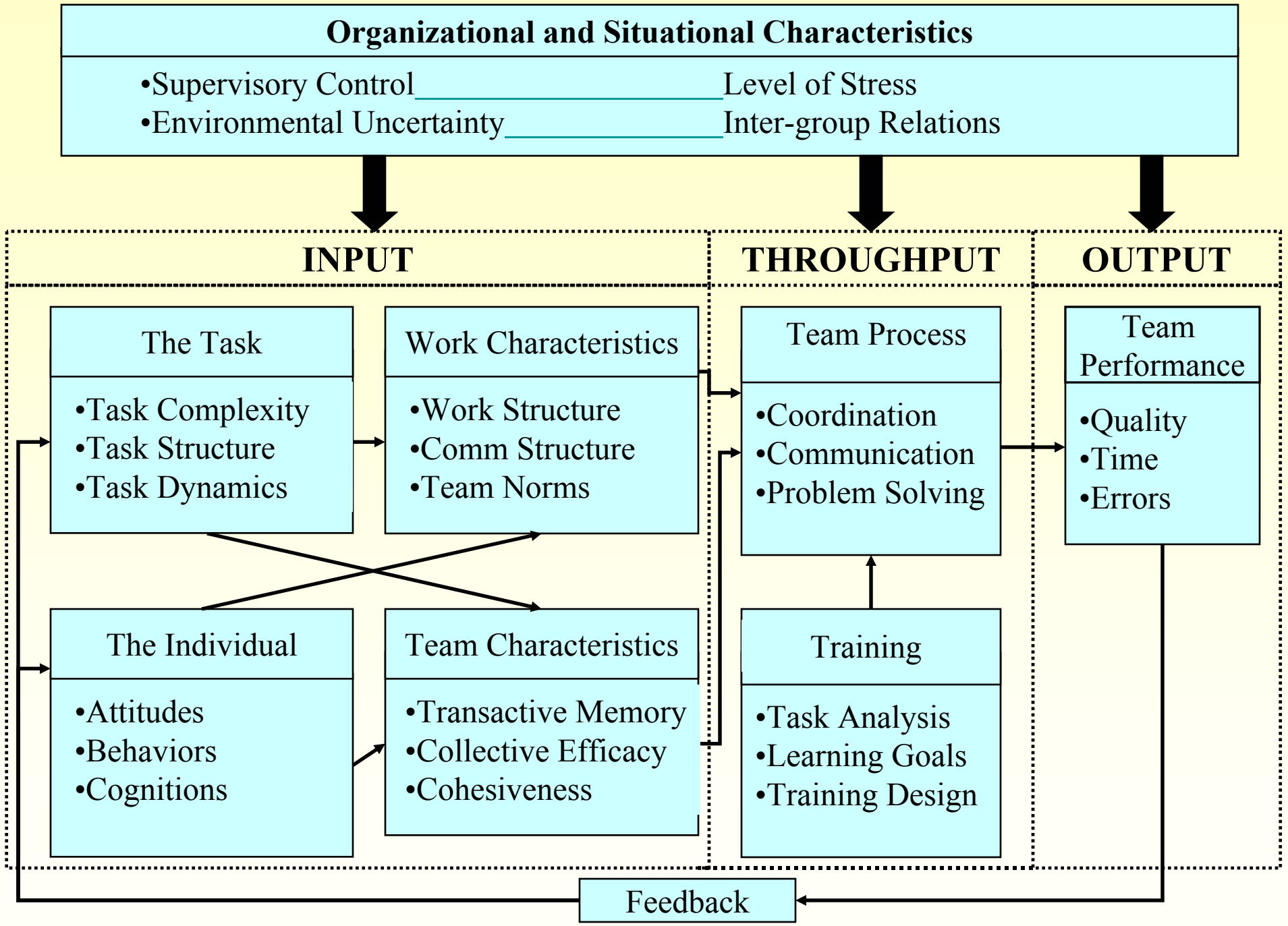
University of Central Florida

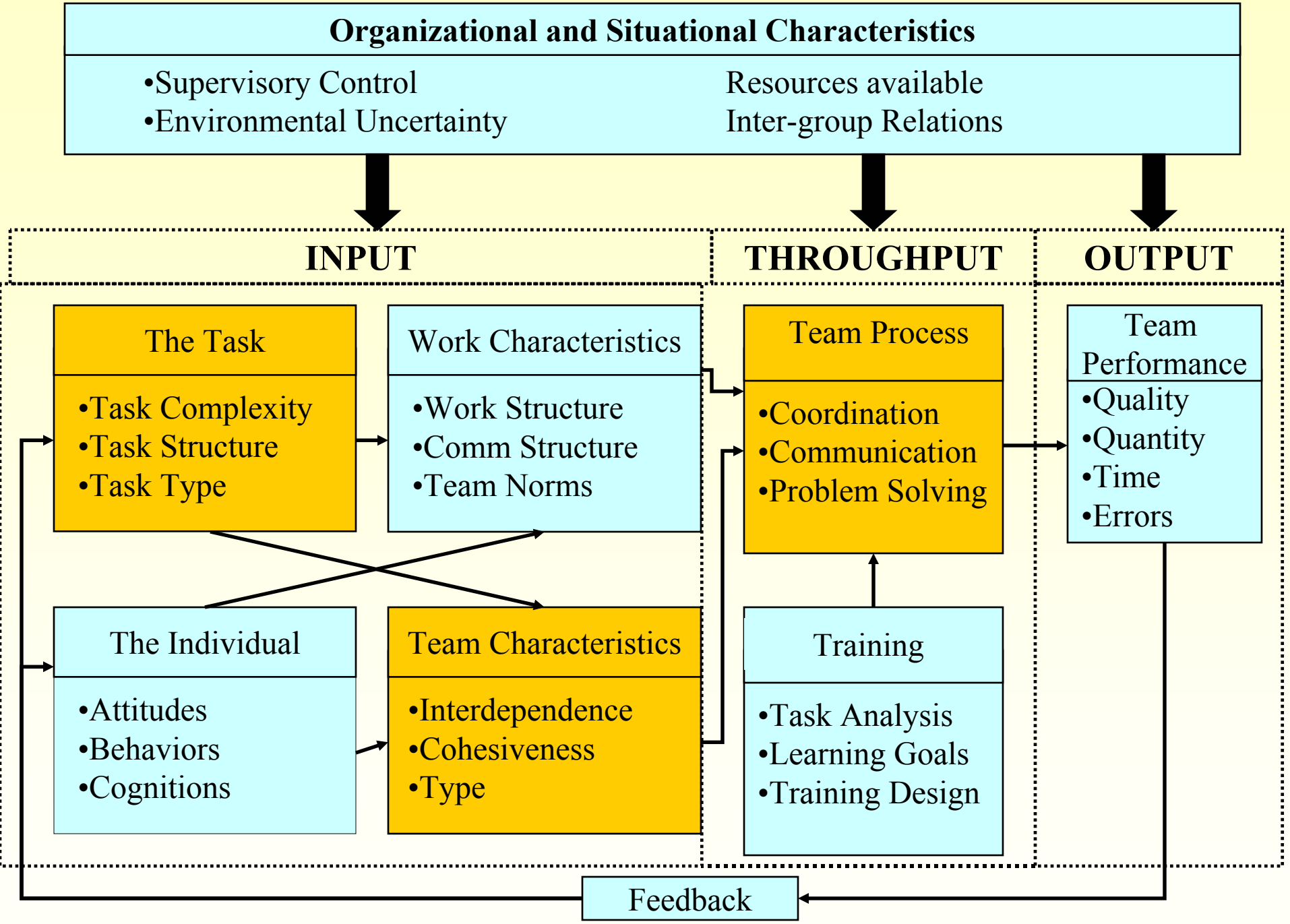
A Word from Our Sponsor

- The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Introduction

- UCF's MURI effort includes a component to see if the results obtained with individuals generalize to teams.
- This component also seeks to identify unique effects of stressors on teams.
- In addition, we seek to identify how different types of teams respond to stressors.


Organizational and Situational Characteristics
•Supervisory Control
Level of Stress
•Environmental Uncertainty
Inter-group Relations
INPUT
THROUGHPUT
OUTPUT
The Task
•Task Complexity
•Task Structure
•Task Dynamics
Work Characteristics
•Work Structure
•Comm Structure
•Team Norms
Team Process
•Coordination
•Communication
•Problem Solving
Team Performance
•Quality
•Time
•Errors
The Individual
•Attitudes
•Behaviors
•Cognitions
Team Characteristics
•Transactive Memory
•Collective Efficacy
•Cohesiveness
Training
•Task Analysis
•Learning Goals
•Training Design
Feedback


Organizational and Situational Characteristics
•Supervisory Control
•Environmental Uncertainty
Resources available
Inter-group Relations
INPUT
THROUGHPUT
OUTPUT
The Task
•Task Complexity
•Task Structure
•Task Type
Work Characteristics
•Work Structure
•Comm Structure
•Team Norms
Team Process
•Coordination
•Communication
•Problem Solving
Team Performance
•Quality
•Quantity
•Time
•Errors
The Individual
•Attitudes
•Behaviors
•Cognitions
Team Characteristics
•Interdependence
•Cohesiveness
•Type
Training
•Task Analysis
•Learning Goals
•Training Design
Feedback

Task Characteristics

Task Characteristics

- Complexity
- Structure
- Type

Stressors:

- Time Pressure
- Work Load
- Fatigue
- Ambiguity
- Noise

Team Characteristics

Types of teams in the literature:

- **Distributed vs Co-located**
  - Virtual and/or real life distribution and location.
- **Interdependence**:
  Reciprocal, Sequential, Pooled, and Team (Saavedra, Earley & Van Dyne, 1993).
- Type of **cues** available:
  -Visual cues of team mates' actions, verbal cues from team mates, etc.
- **Performance feedback**
  - Team or individual.

Team Characteristics

- Team Type
  - Distributed vs Co-located
- Interdependence
- Cohesiveness

Team Processes

Team Processes

- Coordination
- Communication
- Problem Solving

- Explicit Coordination determine coordination strategies under low workload coordination.
- Implicit coordination takes over as work load increases.
  - Serfaty, Entin, Johnston, 1998
- A switch from explicit to implicit coordination is the primary adaptation mechanism that allowed teams to maintain and improve their performance under a high level of time pressure.
  - Entin & Serfaty, 1999

Adaptability

- Adaptability can be viewed as a sub component of communication and coordination and is an important part of effective team process and performance.
- Involves the adjustment of strategies by using information from the environment through compensatory behavior and reallocation of resources in stressful or novel situations. (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1999).

Research Approach

- What the team performance literature includes and what it is lacking.

- Research Questions:

1. **What type of stress effects generalize from individuals to teams and what effects are unique to teams?**
2. **What stress effects generalize across teams and what effects may be unique depending on team type?**
3. **How do teams adapt when circumstances change and how do they behave in response to stressors?**

Team Types

Characteristics	Flashpoint™	DDD	Ghost Recon ™
Task Interdependence Saavedra, Earley, & Van Dyne, 1993	Sequential	Reciprocal	Team
Available cues of team mates' actions	Verbal Only	Verbal and Low Detail Visual	Detailed Visual and Verbal
Similar to:	Reconnaissance Team	Command and Control Team	Special Forces Combat Team

FLASHPOINT™ TESTBED

FLASHPOINT™ OBJECTIVES

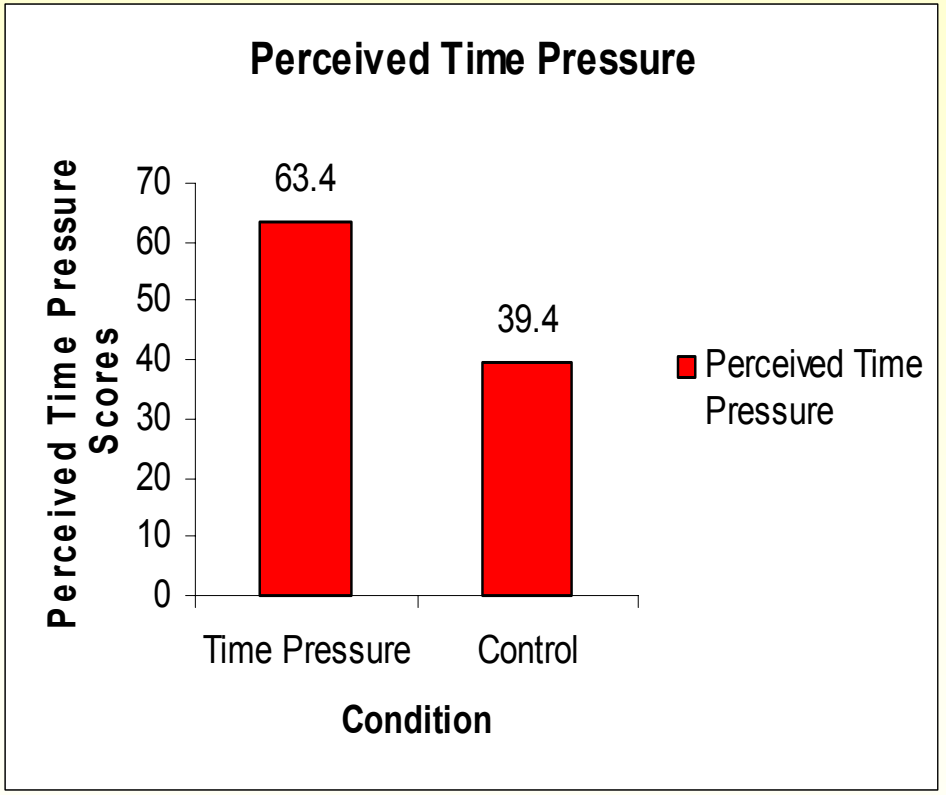
- **Intelligence based** mission with **verbal communication** as the key component to successful performance.
- **Two member distributed action team**
- Objective is to **verify truthfulness** of obtained information regarding terrorist activity by locating objects, hostages and **identify enemy targets** within **3D setting**.
- Teams make a **decision** to release or detain information sources depending on their affiliation (terrorist or allied).



FLASHPOINT™ RESULTS

FLASHPOINT RESULTS:

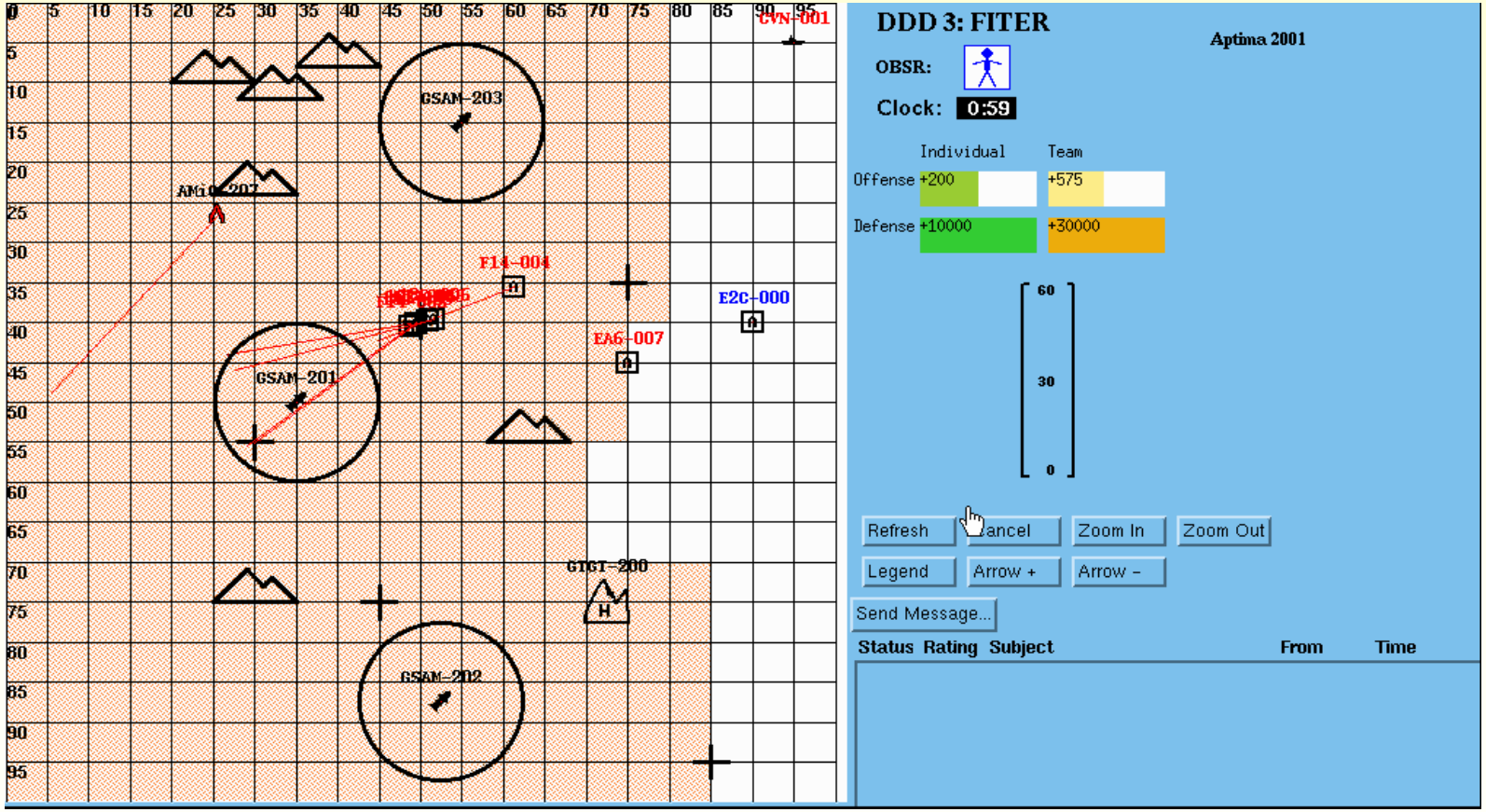
- **3 pilot studies**
- **Manipulation check of perceived time pressure using the NASA TLX has yielded significant group differences.**
- **Preliminary performance data shows a trend toward performance decrements for teams in the time pressure condition.**



DDD TESTBED

DDD OBJECTIVES:

Goal is to fly a **checkpoint mission** and drop bombs on a target at a specified time.



DDD TESTBED

DDD TASK:

Participant: identified unknown aircraft and ground sites and monitored timing and dangerous cue patterns (late to a checkpoint; enemy aircraft approaching the team member); alerted team member of threatening events.

Confederate: flew checkpoint mission, illustrated events via creating situations (e.g. being early or late to a checkpoint; targeting friendly aircraft).

DDD MISSION:

- Scenarios crafted with specific cue patterns to occur throughout the mission.

Dependent Measure

- High workload legs provided more opportunity to perform team performance behaviors.
- To address this, the dependent measure was the proportion of correct **adaptive team performance skills demonstrated** out of the total number of possible.

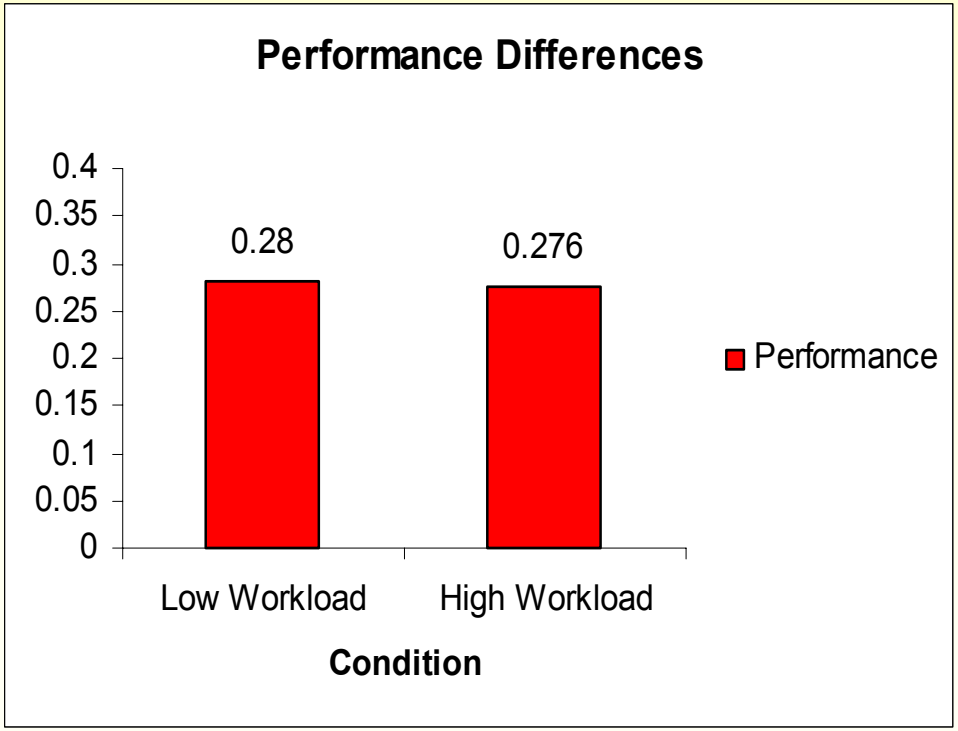
DDD RESULTS

Workload: operationalized as the number of events that occurred during mission legs.

- Low workload: 2-7 events
- High workload: 9-13 events
- Total number of mission legs analyzed: 10

DDD RESULTS:

– No differences in adaptive team performance between high and low workload mission legs.

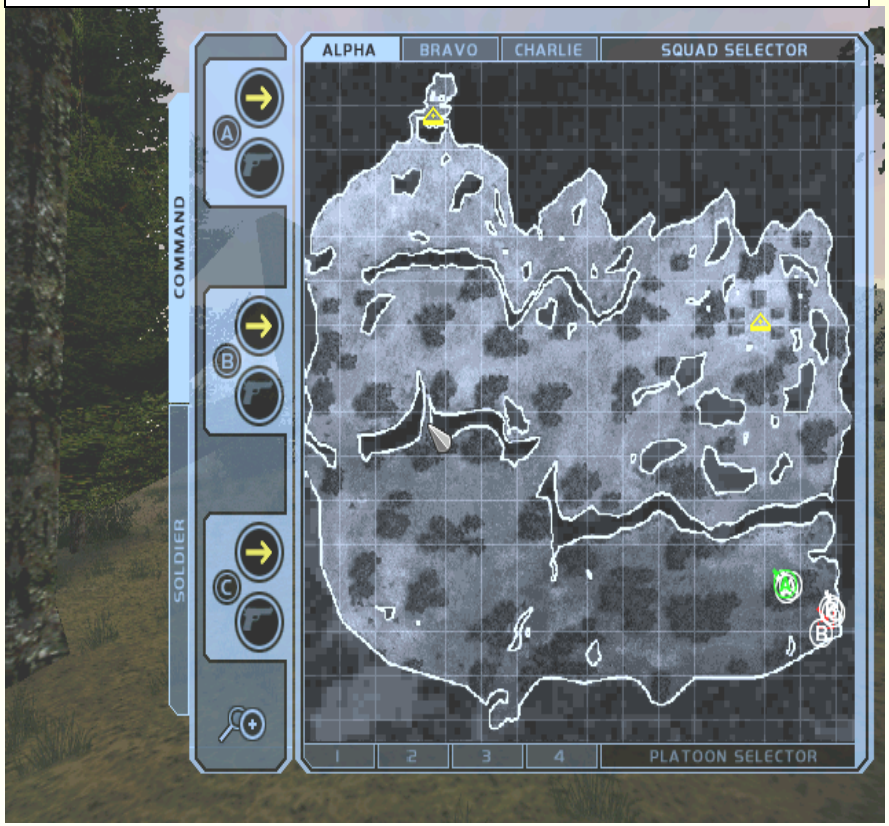


GHOST RECON™ TESTBED

- **GHOST RECON™ MISSION:**
  - Goal is to complete **4 objectives** (e.g., rescue a hostage, secure a building, destroy a SAM site) in a specified time period.
  - Subjects will be randomly assigned to a **three member team** equipped with specific tools (i.e., weapons) necessary to complete the specific mission.
  - **Time pressure and workload**, similar to the methods used in the other 2 test beds, will be used as the stressors.
  - Outcome measures include:
    - Incidents of Fratricide
    - Communication
    - Mission Completion
    - Mission Completion Time

Mission1:

1. Neutralize Tent Camp Troops
2. Secure the Caves
3. Return to Insertion Zone
4. Capture Papashvili



GHOST RECON™ RESULTS

- GHOST RECON™
  - Training scripts, maps, and mission objectives have been finalized.
  - Preliminary pilot studies have yielded observable teamwork behaviors.
  - The NASA TLX will be used as a manipulation check.



Future Research

- Research Questions:
  - Why would there be differential stress effects in teams with different characteristics?
    - What might the effects be? What about their differences influences the stress effects?
  - What are the differences in moderators and coping styles between teams with different characteristics?
    - Are there meta-behaviors that are universal across teams?
  - How do teams shift from operating in one situation to another (e.g., from attack to retreat) in a changing environment?
    - How do teams adapt?
- Objectives: Look at Teams, Adaptability, and Coordination more specifically.
  - Current literature approaches teams and training globally.

DOD Multidisciplinary Research Program: MURI Operator Performance Under Stress (OPUS).

White Paper

On Mental Resources and Performance Under Stress

J.L. Szalma and P.A. Hancock

Department of Psychology and the Institute for Simulation and Training

University of Central Florida

Work Performed under the U.S. Army Research Office: MURI: Optimizing Cognitive Readiness Under Combat Conditions

MURI-OPUS List of Collaborators

Universities:	(Formal Connection): Kansas State University, University of Minnesota, Catholic University, University of Cincinnati. (Informal Connection): Florida Institute of Technology, University of Toronto.
Industry:	United Defense, DCS Corporation.
Government:	NAWC-TSD
Individuals:	Dr. Jerry Krueger

Abstract

With the clear empirical failure of gated stage models of information processing to account for multiple task performance outcomes, Kahneman (1973) offered resource theory as an alternative explanatory construct. Considered primarily as a metaphor, resources have, over the years, remained unidentified as a neurophysiological process or neuroanatomic structure. The failure to establish resources as hypothetical constructs has been interpreted by some as the demise of resource theory itself. Contemporary brain imaging techniques have begun to identify blood flow, glucose transmission, and, potentially, correlates of cortical temperature as resource foundations. The present paper elaborates upon resource theory, and articulates some contemporary proposals on resource differentiation and resource utilization under stress as a foundation for a program of theoretical and experimental progress.

ON MENTAL RESOURCES AND PERFORMANCE UNDER STRESS

Resource Metaphors

Most early resource theories, including Kahneman's (1973) original view and modifications by Norman and Bobrow (1975), Navon and Gopher (1979), and Wickens (1980; 1984), conceptualized them as commodities or as pools of energy to be "spent" on task performance. In general, either economic or thermodynamic/hydraulic metaphors tend to be adopted when describing resources. The economic model is reflected in the description of resources in terms of supply and demand: Performance on one or more tasks suffers when the resource demands of the tasks exceed available supply. Presumably, the total amount of this supply fluctuates with the state of the individual, with the "assets" diminishing with increases in the intensity or duration of stress. Although Kahneman's (1973) original conception allow for dynamic variation in resources, most early models assumed a fixed amount of resources (see Navon & Gopher, 1979). In thermodynamic analogies, resources are a fuel that is consumed, or a tank of liquid to be divided among several tasks, and under stressful conditions the amount of resources available is depleted and performance suffers. In discussing his version of resource theory, Wickens (1984) warned that the hydraulic metaphor should not be taken too literally, but most subsequent descriptions of resources have employed visual representations of resources as just this tank of liquid form. Similarly, many discussions of resource availability and expenditure adopt the economic language of supply and demand, and Navon and Gopher (1979) explicitly adopted principles of microeconomics in developing their approach.

A problem with using non-biological metaphors to represent biological systems is that such models often fail to capture the complexity and the unique dynamic characteristics (e.g. adaptive responses) of living systems. For instance, a hydraulic model of resources links the activity of a tank of liquid, governed by thermodynamic principles, to the action of arousal mechanisms or energy reserves that are allocated for task performance. However, a thermodynamic description of the physiological processes underlying resources is at a level of explanation that may not adequately describe the psychological processes that govern performance. Thermodynamic principles can be applied to the chemical processes that occur within and between neurons, but they may be less useful in describing the behavior of large networks of neurons[1]. Similarly, economic metaphors of supply and demand may not adequately capture the relation between cognitive architecture and energy allocated for their function. Economic models of resources define them as commodities to be "spent" on one or more activities, and they assume an isomorphism between human cognitive activity and economic activity, an assumption which may not be tenable. Indeed, Navon and Gopher (1979) admitted that their 'static' economic metaphor for multiple resources may need to be replaced by a dynamic one that includes temporal factors (e.g. serial versus parallel processing; activity of one processing unit being contingent upon the output of another). Such concerns over the metaphors used to describe resources are hardly new (Navon, 1984; Wickens, 1984), but their use has become sufficiently ingrained in thinking about resources and human performance that reevaluation of the metaphors is warranted. As we shall see, a regulatory model based on physiology may serve as a

[1] *The argument here is not that neural structures are not constrained by the laws of thermodynamics—clearly they are—but that thermodynamic principles implied by the metaphor are not sufficient for the development of a complete description of resources and their relation to cognitive activity.*

better metaphor (and, in future, may serve to describe resources themselves to the extent that they can be established as a hypothetical construct) to describe the role of resources in human cognition and performance. However, even a physiologically-based theory of resources must be tempered by the problems inherent in reducing psychological processes to physiological activity.

Development of a Single Resource Theory

As a precursor to the emergence of resource theory, Moray (1967) argued that Broadbent's (1958) filter theory, and Treisman's (1969) revision, does not apply to all tasks requiring divided or selective attention. Specifically, these models do not adequately account for the performance of well-trained observers on divided attention tasks, or for cases in which the perceptual inputs and the responses (outputs) are compatible. Using the language of Information Theory, he argued that the human should be considered a limited capacity processor rather than a fixed, limited capacity channel. It is the limitation of the central processor that influences performance rather than the perceptual input channels. The mental operations performed on perceptual inputs consume capacity (an aspect of attention not considered by Broadbent or Treisman), and capacity can be divided among different "processors" when divided attention is required. Moray was describing what would later be called "allocation of resources." He also argued that the demands of the task control how this allocation occurs (the capacity allocated to a given channel). Note that in this context capacity is viewed in terms of information processing structures or channels rather than as pools of energy posited by more energetic based models.

An early influential resource theory was that of Kahneman (1973), who argued that there was a single pool of limited capacity available for a variety of tasks, and that performance depends upon the degree to which that capacity is allocated to the task. The amount of capacity available was assumed to be limited and a function of arousal level. An allocation strategy determines how much processing capacity various "activities" (presumably this includes cognitive processes) receive, and the strategy adopted is influenced by characteristics of the individual ("enduring dispositions") and motivational influences ("momentary intentions").

Kahneman's (1973) theory was extended by Norman and Bobrow (1975), who are credited with coining the term "resources" in the context of attention and human performance. They distinguished resource-limited from data-limited processes. Resource-limited processes are those that improve as resource investment is increased, and data-limited processes are those in which no further increase in resource allocation will impact performance, which is driven by the quality of the information (data) available to the individual. According to Norman and Bobrow, there are cases in which performance is resource independent; this has implications for optimal resource allocation. The effectiveness of allocation depends not only on environmental demands but also on limitations imposed by the structure of the task. The distinction between resource and data-limited processes can account for "flat" portions of the performance-resource function, in which performance is unaffected by increases in resource allocation or effort. Note, however, that in some situations an apparent data-limited process may actually be a simple ceiling effect created by the level of task difficulty chosen by the experimenter (Kantowitz & Knight, 1976), although such a conclusion is somewhat controversial (see Norman & Bobrow, 1976).

The data-limited process concept was expanded by Navon and Gopher (1979) who used the term subject-task parameters, which include the task characteristics (quality of the data), the environment, and the permanent properties and transient states of the subject. Hence, their view

extended data-limited processes to include task-operator interactions, a conceptualization consistent with transactional models of stress (Lazarus & Folkman, 1984; Matthews, 2001). The subject-task parameters define the constraints within which resources can be recruited and allocated for task performance. In addition, they noted that rather than being a fixed capacity system, resources may be somewhat flexible, fluctuating with arousal, such that in some situations increasing task load may increase arousal and release a larger supply of resources.

Multiple Resources

An alternative to a single resource view is multiple resource theory. Navon and Gopher (1979) argued that the human is a multiple channel processor (multiple structures), that each processor may have its own capacities, and that each capacity may be shared by several processes. Hence, the impact of changes in task difficulty on resources may not be purely quantitative, but may also be qualitative (i.e. structural). Different subject-task parameters may differentially influence the resource demands on diverse processing structures, and specific parameters may influence the amount of resources demanded for a particular structure. A question to be addressed is whether different stressors differentially affect specific information processing structures available, or whether the effects of stressors are more global in character (i.e. affecting all structures simultaneously, either to the same or to different degrees). Navon and Gopher only mentioned stress in the context of arousal and resources, indicating that high arousal accompanying high task load is a state of stress rather than of processing capacity. However, this brief treatment implies an arousal model of stress without considering the multi-dimensionality of the stress construct or the impact of stress on cognitive structures and the quality of information processing (see Matthews, 2001).

In an influential version of multiple resource theory, Wickens (1984) proposed separate pools of resources for spatial and verbal information processing codes, different modalities,[2] stages of information processing (encoding, central processing, and responding), and motor vs. verbal response. These resources can be utilized independently or jointly depending on the demands of the information processing components that comprise a task.

Multiple resource views developed as a result of experimental findings not easily accounted for by single resource views. For instance, Wickens (1984) argued that single-resource theories cannot readily explain perfect time-sharing between tasks, nor can they explain why changes in task structure that leave difficulty constant (and, presumably, the amount of resources required for task performance) lead to a change in the degree of interference with a secondary task. In addition, Wickens described two aspects of task difficulty that cannot easily be accounted for by single resource views: 1) Difficulty insensitivity; increases in the demands imposed by a primary task do not influence performance on a secondary task. 2) Uncoupling of difficulty and structure; cases in which two tasks are each paired with a third task, but the more difficult of the two tasks interferes less with the third task than the easier one.

Wickens argued that a multiple resource model can account for each of these issues. Perfect time sharing can occur when there is no overlap in resource demand between two tasks, as

[2] *The original model specified only audition and vision, but in principle the model would apply to other modalities as well.*

would be the case with an auditory shadowing task and a secondary tracking task. Cases in which task structure changes do not influence the overall amount of resources required (shifting from an auditory to a visual task) but impair secondary task performance is due to a shift in resource demands from separate pools to a common reservoir of resources. Difficulty insensitivity is accounted for by cases where the two tasks impose demands on different pools of resources, and uncoupling of difficulty and structure occurs when the more difficult task shares resources with the third task but the easier task does not. In arguing for the existence of multiple resources, Wickens (1984) was careful to point out that this view does not eliminate the possibility of an additional set of general resources available for use in any task.

Definition of Resources

A major problem for resource theory, whether single or multiple, is defining the concept itself. Wickens (1984) defined resources as an intervening variable to account for variability in the efficiency with which individuals can divide attention among tasks. He made the distinction between resources as something available for use (a "commodity"), versus resources as distinct information processing structures, in which the degree to which two tasks interfere with one another is determined by the degree to which they compete for common structures (e.g. spatial, verbal). He also argued (Wickens, 1992) that capacity, which represents the maximum or upper limit of processing capability, should not be confused with resources, which refer to scarce supplies of mental effort allocated to improve processing efficiency. Alternatively, Gopher (1986) argued that structure and energy (or the "commodity" in Wickens' terms) are distinct components of resources. Indeed, the multiple resources concept inherently includes a structural element. However, this distinction is blurred by ambiguous terms like "processing resources," which can be interpreted as energy or structure (or both). This has contributed to the confusion surrounding the concept.

There have been attempts to identify physical structures that serve as resources. Wickens (1984) discussed several candidates, which if successful would allow the concept of resources to shift from its status as an intervening variable to that of a hypothetical construct. He suggested that pupil diameter may be a useful index of resources, as it is a manifestation of ARAS activity. However, such a view emphasizes resources as energy reserves and neglects the importance of structure inherent in a multiple resource view. To support the latter view, Wickens noted that there is some evidence for "functional cerebral space," for which different tasks compete and which may ultimately serve to define resources as neural processing mechanisms (see Kinsbourne & Hicks, 1980). From this perspective, there is resource competition to the degree that task demands share the same neural structures. This implies that resources are processing capacity in the computer sense, where the processing chip and working memory of the computer are used as a metaphor for psychological processing (a third metaphor applied to resource theory more recently, although see Moray, 1967).

It is also possible that the cognitive structures (information processors) are the resources. Rugg (1986) suggested that the need for the concept of resources arises from the current inability to adequately specify the cognitive architecture responsible for information processing. If the specific components of information processing structures that drive performance could be articulated, then the number of resources would be the same as the number of active components; the resource model thus becomes unnecessary (see also Navon, 1984). Resource

limitations reduce to structural limitations when the structures are adequately specified. However, such a view, while acknowledging the importance of cognitive structure to resource theory, does not reflect the importance of the energetic component of behavior. In addition, Rugg himself argued that the link between resources and physiology was not clear, since the physiological changes could be a consequence of the mobilization of resources (the body's attempt to adapt to the behavioral demands).

Resources: Structure and Energy. As noted earlier, Gopher (1986) argued that resources consist of structural and energetic components. Specifically, resources are aggregations of elementary units defined by the type of processing activity they serve and the range and rate of activation. The degree of activation of those units is determined by the energetic component. Gopher noted that most views of resources emphasized the energetic over the structural elements, an observation that is true sixteen years later. He acknowledged, however, that Kahneman's (1973) original model did include structural components in the form of the possible activities to which available capacity (resources) was allocated. Wicken's multiple resource model also contains structure, inherent in the definitions of the qualitatively different pools of resources (e.g. stages of processing, coding, etc). Gopher's view of structure, however, goes beyond these general conceptions to consider specific units of information processors that comprise, in his model, a vector for structure that combines with energetic sources for each processor, the product being system response. Hence, in Gopher's model, the response of the overall information processing system is a product of two vectors: structure and energy.

If resources consist of both cognitive structure and energy, then the concepts of resource investment and scarcity of resources have multiple interpretations, which has likely contributed to confusion in the development and tests of the theory. Gopher (1986) addressed the different manifestations of both resource investment and scarcity. He described three forms of modulation in intensity of resource investment: 1) changing the number of processing units engaged in a task, 2) changing the activation level of those units, or 3) changing the time each processing unit spends on a task. The first and second forms of resource investment are the most often employed in applications of resource theory to research on stress and performance, a bias that may be due in part to the history of the theory as an answer to the problems of arousal models in explaining the effects of stress on performance. This has led to the neglect of the relation of the third form of investment to stress (although see Hancock and Weaver, in press, for a noteworthy exception).

While scarcity of resources can be described in terms of the amount available for consumption, Gopher (1986) discussed another aspect of scarcity that is consistent with a physiological analogy: the rate at which they can utilize resources. This represents the structural limitations of the processing units. For instance, if a car has a full tank of gas, but the flow of fuel to the engine is maximized, no increase in the pool of resources will help, even though performance appears "resource limited." Here the scarcity in resources is not in their amount, but in their rate of usability. He also noted that time and whether resource scarcity is transient or steady state can also influence the nature of resource limitations. Note that this view of resources and its implications for investment and scarcity is beyond the relatively simplistic economic and thermodynamic metaphors often applied in descriptions of the theory.

A Physiological Model

Recently, Cabanac and Russek (2000) outlined a representation of regulated biological systems that may be useful for conceptualizing the energy component of resource theory. They adopted characteristics of control theory and applied them to biological systems. Physiological systems reach a steady state, such that a constant amount of free energy available for use is maintained and the inflow and outflow of energy are equal and constant. These systems have a set point that indicates the normal level of function. Perturbations of this steady state require the system to compensate for deviations from the set point (this is essentially a homeostatic model). Further, Cabanac and Russek (2000) argued that the regulation of inflow is anatomically distinct from the regulation of outflow. The inflow regulation is negative feedback loop in which a standard reference (presumably determined by the set point) is compared to current state and an input subsystem responds by increasing flow when energy is drained and decreasing it when energy levels rise above the set point (a negative loop). The outflow is a positive (feedforward) loop in which changes in the state relative to the set point induce changes in the same direction in the output subsystem (e.g. increase in amount of energy induces increase in outflow).

They also distinguished between regulated variables, those that are maintained at a constant level, and controlled variables, which correct the deviations of the regulated variables. They delineated three properties of any system which continuously modulates the inflow/outflow of energy or mass: 1) the minimum flow rate through the system, which is the set point for the regulated variable. Perturbations change either inflow or outflow, which changes the total flow rate. 2) The maximum continuous perturbation the system can regulate against, which is equal to the maximum flow rate of the system [this presumably is limited by the structure of the system—e.g. the neuro architecture and function limitations; at higher levels, the processing capacity of attention and cognition]. 3) The speed at which the system can return to set point after transient perturbation is proportional to the magnitude of the baseline flow rate.

Although Cabanac and Russek (2000) argued that inflow and outflow processes were anatomically distinct, it may be difficult to empirically distinguish them at the level of perception and cognition (and they may not be anatomically distinct when applied to the nervous system). If one applies their model to psychological processes, the "tank" of energy becomes a pool of resources (as energy) which are continuously drained and replenished, thus creating a steady state. Perturbations would result in deviations from set point (steady state) and require regulatory response in the form of either increasing/decreasing inflow or outflow, and the depletion of resources would occur when outflow is greater than inflow for a specified period of time. The regulatory response would presumably be controlled by the cognitive structures responsible for resource allocation (e.g. a "central executive;" see Baddeley, 2001). Stressors, as a perturbation, require increasing the inflow, since outflow cannot be decreased without sacrificing task performance. Assuming a motivated operator, performance declines as the ability to replenish diminishes. Alternatively, stress could increase outflow, as more processors are engaged and more energy expended to cope with the stress while maintaining performance. In either case, a threshold (for the amount of available resources) may exist below which performance begins to suffer. Note that an implication of the feeedforward loop is that one can have "too many" resources, such that the amount of resources available exceeds steady state levels. In such situations, a person would seek activities to consume those resources and performance might suffer as a result of understimulation (see Hancock & Warm, 1989).

This model addresses the energetic component of resources, but it does not address structural considerations, beyond the maximum perturbation a system can handle or how resources are allocated for information processing, both of which are presumably constrained by structural factors. The maximum amount of energy that can be applied to compensate for perturbation will, if we apply Cabanac and Russek's (2000) model, be equal to the maximum flow rate of the system, presumably a structural feature. This may represent a point at which a process becomes data-limited. The characteristics of the information processors, however, remain unspecified by the physiological model. An additional structural concern is whether the quality of the information processing degrades (independent of energy level) with increases in the intensity and/or duration of stress.

Issues Surrounding the Resource Concept

There are several issues regarding resource theory that need to be addressed if it is to be useful for advancing stress theory and clarifying the link between stress and performance. These concerns are addressed in this section.

Single vs. Multiple Resources. It is possible that rather than a set of qualitatively different resources, there may be a single resource that can be more easily ("naturally") divided for particular sets of activities. An analogy from classical conditioning may clarify the issue. Certain S-R relations are learned faster than others. For instance, learning to avoid a stimulus after it is associated with nausea can occur in one trial, while learning to avoid the same stimulus when it is paired with shock is learned more slowly (Rozin & Kalat, 1971). Similarly, it may be that processing resources are more easily separable for some sets of tasks than for others. Indeed, Moray (1967) noted that selection of attention is easier along some dimensions than others. Hence, the differences in time sharing performance may have to do with ease of allocation of a single resource rather than multiple resources per se. However, these two concepts may be empirically indistinguishable. Ultimately, the issue reduces to the more fundamental problem regarding the identity of resources, as discussed later in this section.

Function of Resources. Another problem for resource theory is the absence of a precise description of how resources control different forms of information processing. Do resources determine the energy allocated to an information processor, do they provide the space within which the processing structure works, or does the processor draw on the resources as needed (and available). In the latter case, the cognitive architecture would drive energy consumption and allocation, but the locus of control for the division of resources remains unspecified in any case. Presumably, an "executive" function that either co-ordinates information processors drawing on different pools of resources or decides how resources will be allocated must itself consume resources, in terms of both energy required for decision making and mental "space" or structure required. Hence, resource theory does not solve the homunculus problem for theories of attention, nor does it adequately describe resource allocation strategies behind performance of information processing tasks.

Empirical Tests of the Model. Navon and Gopher (1979) commented on the problem of empirically distinguishing declines in performance due to insufficient supply from those resulting from increases in demand. They noted that, "When the performance of a task deteriorates, is it because the task now gets fewer resources or because it now requires more?" (p 243) This is similar to the issue concerning whether resource depletion results from changes of

inflow or outflow discussed earlier in the context of Cabanac and Russek's (2000) physiological model. Navon and Gopher (1979) characterized the problem as distinguishing between changes in resources and changes in the subject-task parameters that constrain resource utilization, and they offered two approaches to avoid this difficulty. One approach is to define the fixed constraints of the task and observe how the information processing system manages the processes within those constraints. The degree of freedom of the system, in this view, is the pool of resources available, in which the term 'resource' is interpreted broadly to include quality of information, number of extracted features, or visual resolution. The subject-task parameters define what is imposed on the system (the demands) and the resources refer to what the system does in response to the demands (allocation of processing units). From this perspective resources can be manipulated by the information processing system within the constraints set by the subject-task parameters. A second approach is to distinguish the kind of control the system exerts on resources, between control on the use of processing devices (what we have called "structure") and the control of the properties of the inputs that go into these devices. The 'devices' are processing resources. The other kind of control is exerted on input resources, which represents the flexibility the person has for determining which inputs are operated on, as determined by subject-task parameters. Processing resources are limited by the capacities of the information processors, while the input resources are limited by subject-task parameters (and allocation strategies that determine which information the operator attends to). Presumably, the individual would have some control over the allocation strategy, in terms of the processing resources devoted to a task, although these can also be driven by task demands (e.g. a spatial task requires spatial processing units). Navon and Gopher did not advocate either approach, but presented them as alternatives for further investigation. The implication for examining the resource model of stress is that one must manipulate both the subject-task parameters (e.g. by varying the psychophysical properties of the stimulus, manipulating the state of the observer, or varying the kind of information processing demanded by the task) as well as the allocation strategies the operator uses (the input resources—e.g. payoff matrices, task instructions). This would provide information regarding how specific stressors impair specific information processing units and how they change operator resource allocation strategies in the presence of continuous stress.

In a later article, Navon (1984) moved to a position less favorable toward resources than the earlier approach (Navon & Gopher, 1979). In the later paper, Navon asserted that predictions derived by resource theory could be made, and results explained, without appealing to the resource concept (see also Rugg, 1986). One could instead interpret effects in terms of the outputs of information processors. Most manipulations, such as difficulty (which in his view influences the efficiency of a unit of "resources") or complexity (which affects the load, or the number of operations required) influence the demand for processing, with supply having no impact upon their interaction. However, this approach assumes a clear distinction between outputs of a processing system and the concept of a resource, and Navon's (1984) notion of specific processors seems blurred with the notion of a resource, as both are utilized for task performance. Nevertheless, his critique that the resource concept is too vague is relevant, and Navon did argue that if resources are viewed as an intervening variable rather than a hypothetical construct, the concept has utility for describing the process. According to Navon, it doesn't matter whether the economic terminology is meant literally or metaphorically, although, as noted earlier, metaphors taken too literally can be misleading.

Structural Mechanisms. If different kinds of information processing draw on different kinds of resources, in terms of the information processors engaged in a task, stressors may have

characteristic effects on each "resource." In addition, as Navon and Gopher (1979) have noted, an aspect of resource utilization is the efficiency of each resource unit. It may be that stress degrades the efficiency of information processing units, independent of energy level or allocation strategy. Investigation of such effects could be accomplished by transitioning between tasks requiring different kinds of information processing and determining if the effects of stress on one structure impacts the efficiency of a second structure.

The quality of resources can vary not only in terms of the kind of information processing unit engaged, but also in terms of the kind of task required. Following Rasmussen's (1983) classification system for behavior, some tasks require knowledge-based processing, in which the operator must consciously rely on his/her mental model of the system in order to achieve successful performance. Other tasks fall under the category of rule-based behavior, in which a set rules or procedures define task performance. The third category is skill-based behavior, in which the task is performed with a high degree of automaticity. Presumably, each kind of task requires different amounts of resources, but they may also represent qualitatively different forms of resource utilization. In other words, these tasks may differ in the efficiency of a unit of resources. As task performance moves from knowledge to rule to skill based processing (e.g. with training), the cognitive architecture may change such that fewer energy reserves are required, and those that are engaged become more efficient. Moreover, the way in which each of these systems degrade with time under stress may be systematic, with the more fragile knowledge-based processing degrading first, followed by rule based processing, with skill based processing degrading last (at this point, one may begin to see breakdown of not only psychological processes but physiological ones as well—see Hancock & Warm, 1989). This degradation may follow a hysteresis function, such that a precipitous decline in performance occurs as the operator's resource capacity is reduced below a minimum threshold for performance. Moreover, these processes may recover in an inverse form, with skill-based processing recovering first, followed by rule and knowledge-based processing. However, based upon the physiological model outlined earlier, one would expect that expanding available resources (opening up the inflow valve) would allow more resources to be applied to tasks and would have a greater impact upon knowledge rather than rule or skill based processing.

Energetic Mechanisms. To investigate the energetic aspects of resources, one must manipulate environmentally-based perturbations, in the form of external stressors (noise, heat) and task demands, to systematically affect inflow versus outflow of energy. Presumably, inflow is controlled by arousal levels, physiological energy reserves, and effort. One could examine performance under manipulations of energetic resources under dual task performance (e.g. what happens to performance on two tasks under sleep deprivation or caffeine consumption?). For example, the steady state can be perturbed by increasing (e.g. caffeine) or decreasing (e.g. sleep deprivation) energy while systematically varying the demands for two tasks.

Time is an important component of the process of resource recruitment and allocation, as changes in mental resources occur as a function of time on task. Thus, transitions from one task demand level to another could be investigated in a systematic fashion under conditions of high/low stress. The key challenge is determining whether stressors such as noise, heat, or sleep deprivation slows inflow, increases outflow, or whether they have interactive effects. We must assume that outflow is manifested by behavior at multiple levels (from cellular up to molar behaviors and cognitive processing). Hence, to increase outflow, one could manipulate task demands, in terms of complexity or difficulty. The Hancock and Warm (1989) model applies to

this situation, as increasing effort to cope with stress progressively impairs psychological and eventually physiological functioning. The interaction between inflow and outflow is manifested in the interaction between effort and co-ordination of task demands. It is an open question whether the allocation of resources occurs at the outflow point (multiple spigots), or at the input (multiple resources fed by a central pool), an issue that may not be easily amenable to empirical testing.

Structure & Energy. Another empirical challenge is to distinguish structure from energy. Given the definitional problems associated with the resource concept, it is not clear whether performance declines because of reduction in energy level or degradation in structures (i.e. failures or declines in the efficiency of the processing units), or a combination of both. If structure and energy are distinct elements of resources, it is hypothetically possible to manipulate one while holding the other constant, although the validity of that assumption is questionable. Is it possible to manipulate specific forms of information processing under constant energy level? Is it possible to manipulate energy level independent of which cognitive processes are utilized? If the decline in available resources is, at least in part, due to the degradation of particular information processing units, then transferring to a task requiring the same processor should lead to worse performance than transferring to one that is different. For instance, if a person engages in a task requiring verbal working memory while under stress, then transition to a task requiring spatial discrimination, performance on the latter should depend only on energetic factors, not on structural ones. Note, however, that in this case the effects of different mental capacities would be confounded with the effects of novelty and motivation on performance.

Summary

Resource theory was developed as a response to the limitations in the ability of alternate information processing models, such as bottleneck theories of attention and the arousal model of performance, to account for results on performance of tasks requiring selective and divided attention. Resources are intervening variables thought to control human performance and operator response to workload and stress. Some have argued that resources may be a hypothetical construct, but strong evidence for this has not been forthcoming, mainly because it is difficult to establish a single biological structure or set of structures that correspond to resources. Both economic and energetic metaphors have been used in describing the structure and function of mental resources. This review of resource theory described resource models of attention and human performance, and modifications of those models. The limitations of resource theory were addressed, and the implications of both the theory and its limitations for stress theory were discussed. Specific research questions based upon these concerns are listed below.

Identified Research Questions

- Can a regulatory model based upon physiological processes be applied as a metaphor for resources? Is the energetic element of resources controlled by such a mechanism?

- Can the structural component of resources (information processing units) be specified beyond the currently gross level?

- In addition to resource depletion, there may be strategy changes that occur with time on task. This relates to the differential quality of resources.
    - Do knowledge, rule, and skill based processing degrade progressively?
    - After a stressor is removed, do these three processes recover in reverse order?

- Are the processes of depletion and replenishment of resources empirically distinguishable?
    - If so, is resource scarcity a consequence of depletion over time, a slowing in replenishment processes over time, or an interaction?
    - If energy reserves are manipulated (caffeine, fatigue) and task demands are manipulated (drawing more resources), can relative rates of "inflow" and "outflow" of resources be explored?

- Are there thresholds for resource depletion?
    - If so, do these thresholds vary according to cognitive architecture (the specific forms of information processing employed for a task)?

- Does threshold level relate to hysteresis, such that the precipitous decline in performance occurs as an operator's resources decrease below threshold, and recovery occurs as resources are replenished to threshold levels?

- If resources consist of structural and energetic components, how does stress impact each one? How does stress influence the interaction of these components?

- Do factors that degrade energy levels, such as sleep deprivation or temperature fluctuations, impact knowledge based processing more severely than the other forms of processing?

Anticipated Outcomes

- Increased understanding of how stress impacts information processing and mental resource utilization will promote a deeper understanding of the mechanisms by which stress impacts performance.

- Increased understanding of the relation between resources and stress will facilitate development of theoretical models of stress and performance.

- Identification of factors (task, environment, person) that control resource allocation and consumption.

- A description of the implications of advances in resource theory and research for the design of stress-adapted human-machine interfaces.

References

Baddeley, A.D. (2001). Is working memory still working? *American Psychologist*, **56**, 851-864.

Broadbent, D.E. (1958). *Perception and communication*. London: Pergamon Press.

Cabanac, M. & Russek, M. (2000). Regulated biological systems. *Journal of Biological Systems*, **8**, 141-149.

Gopher, D. (1986). In defense of resources: On structures, energies, pools, and the allocation of attention. In G.R.J Hockey, A.W.K. Gaillard, & M.G.H. Coles (Eds.), *Energetics and human information processing*. (pp. 353-371). Dordrecht: Martinus Nijhoff Publishers.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, **31**, 519-537.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kantowitz, B.H., & Knight, J.L. (1976). On experimenter-limited processes. *Psychological Review*, **6**, 502-507.

Kinsbourne, M., & Hicks, R.E. (1980). Functional cerebral space: A model for outflow, transfer, and interference effects in human performance: A tutorial review. In R. Nickerson (Ed.), *Attention and performance VIII*. (pp. 345-362). Hillsdale, NJ: Erlbaum.

Lazarus, R., & Folkman, S. (1984). Stress, appraisal, and coping. New York: Springer Verlag.

Matthews, G. (2001). Levels of transaction: A cognitive science framework for operator stress. In P.A. Hancock & P.A. Desmond (Eds.), *Stress, workload, and fatigue*. Mahwah, NJ: Erlbaum.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica*, **27**, 84-92.

Navon, D. (1984). Resources—A theoretical soupstone? *Psychological Review*, **91**, 216-234.

Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. *Psychological Review*, **86**, 214-255.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processing. *Journal of Cognitive Psychology*, **7**, 44-60.

Norman, D. & Bobrow, D. (1976). On the analysis of performance operating characteristics. *Psychological Review*, **6**, 508-510.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, 257-266.

Rozin, P., & Kalat, J.W. (1971). Specific hungers and poison avoidance as adaptive specializations of learning. *Psychological Review*, **78**, 459-486.

Rugg, M.D. (1986). Constraints on cognitive performance: Some problems with and alternatives to resource theory. In G.R.J Hockey, A.W.K. Gaillard, & M.G.H. Coles (Eds.), *Energetics and human information processing*. (pp. 353-371). Dordrecht: Martinus Nijhoff Publishers.

Treisman, A.M. (1969). Strategies and models of selective attention. *Psychological Review*, **76**, 282-292.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII.* (pp.239-257). Hillsdale, NJ: Erlbaum.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & D.R. Davies (Eds.), *Varieties of attention*. (pp. 63-102). New York, NY: Academic Press.

Wickens, C.D. (1992). *Engineering psychology and human performance, 2nd Ed*. New York, NY: Harper Collins.

The Triples Rule

Robert R. Hoffman, Patrick Hayes, and Kenneth M. Ford, *University of West Florida*
Peter Hancock, *University of Central Florida*

A fundamental stance taken in human-centered computing is that information-processing devices must be thought of in systems terms. At first blush, this seems self-evident. However, the notion has a long history, and not just in systems engineering.

The Industrial Fatigue Research Board

During World War I, the British Health and Munitions Workers' Committee presented the results of wartime studies of fatigue and efficiency, which motivated the establishment of the Industrial Fatigue Research Board. The IFRB's mandate—marking the beginnings of ergonomics in England—was to identify the causes of worker fatigue and boredom (due to mechanization, task automation, and the need for concentrated attention) and determine ways to alleviate it.[1]

Under the IFRB's aegis and the leadership of pioneer industrial psychologists, studies were conducted of diverse jobs, covering worker selection, product design, production procedures, delivery methods, environmental effects on productivity and safety, and so on.[2] In the pages of the Board's research reports—there were 90 in all, published up through 1947—we find the first use of the phrase "the human factor."[3] We also find the first explicit notion of a man–machine system. In his 1923 "personal contribution" to the IFRB's annual report, H.G. Weston foresaw a need for a program of psychological research:

> The introduction and development of power-driven machines has effected an enormous savings of time and energy, not only by increasing the rapidity of production through substitution of mechanical power for human effort, but also by changing the character of the manipulations which remain to be performed by the operative. So great has this economy been, that is has brought with it a tendency to overlook the possibility that, while industrial machinery may be admirably adapted to the performance of its mechanical functions, it may be incompletely adapted to the needs of the human organism, upon whose efficient co-operation it depends for its productive use.[4]

Weston goes on to mention research detailing the discovery of serious design flaws in various machines that led to inefficiency and physical debilitation. For example, even commonly used lever shapes could result in serious physical debilitation if they were in an improper position. Sometimes, attempts to correct one obvious problem led to another one that the design engineers did not recognize. For example, the lightening of a roller in a laundry machine was not accompanied by a lightening of the load needed to depress the foot pedal controlling the roller, making the roller mechanism more, rather than less, difficult to use.

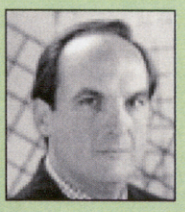
> It is, therefore, most important that correct design should be secured in the first place. It is difficult to see how this can be obtained to the fullest extent except as the result of definite research, undertaken with the object of determining such physiological and psychological facts as should be borne in mind when designing machines, the forms of mechanism and mechanical combinations which will conform to the needs of the operative.[4]

We could therefore say that the basic concept of the man–machine system was already entrenched in industrial psychology as early as the 1930s. To say now that we must think of humans and machines in systems terms seems perhaps more than passé. However, technology has wrought many changes since the 1930s, entailing entirely new implications for the meaning of human–machine system.

Cognitive fit: The dawn of a new age

Computers do things with symbols that reflect the meanings of those symbols, forcing us to rethink the relationship between machines and humans—the cognitive fit as well





Editors: Robert R. Hoffman, Patrick J. Hayes, and Kenneth M. Ford
Institute for Human and Machine Cognition, University of West Florida
rhoffman@ai.uwf.edu

1094-7167/02/$17.00 © 2002 IEEE

as the physical fit and how the two relate. For example, eyeglasses are an ocular prosthesis that improves vision. They must be fitted to the individual's head shape and vision; my eyeglasses won't work well for most other people. However, the use of the ocular prosthesis depends on context: If I am wearing my reading glasses and suddenly look up to examine something far away, the eyeglasses become a hindrance rather than a help. Extending this analogy to computers, we must be aware of contexts in which the human and machine collaborate and operate. This broadens the notion of context and makes it subtler; it now includes such things as goals and expectations, which are just as relevant as physical context. Also, the context can include other people in a way it hadn't before because the machine can facilitate new kinds of communication, and other peoples' concepts and ideas can be part of the context of use.

SRI International scientists in a mobile-agents research project are developing eyeglasses that do not magnify in the usual sense (see www.erg.sri.com/projects/sae/cars-poster.jpg). Rather, they can calculate where you're looking and project a virtual image onto the real scene—for example, to show you where to turn to get to the nearest convenience store. When such devices become commonplace, we will regard them as an extension of our vision and will wonder how we ever got along without them. To not have them—a failure to be able to see in new ways—will be regarded as a form of shortsightedness. However, to make such devices commonplace, we must totally reconceive what it means for a machine to have an interface.

Modern cognitive engineering has progressed through four "ages" over its 30-plus-year lifetime. During the Age of the Average Man, design was based on "one size fits all." During the Age of Adaptation, there arose a recognition of the need for flexibility. During the Age of Personalization, we realized that machine siblings could be adapted to single users. We are now entering a fourth age.

The Triples rule

In this new Age of Symbiosis, machines are made for specific humans for use in specific contexts. The unit of analysis for cognitive engineering and computer science is a triple: person, machine, and context (see Figure 1).

The Triples rule asserts that system development must take this triple as the unit of analysis, which has strong implications, including a mandate that the engineering of complex systems include detailed *cognitive work analysis*.[5–7] It also has implications for the meaning of intelligence, including artificial intelligence.

Intelligent...in a sense

It is easy, perhaps too easy, to regard humans as intelligent, but we must keep in mind that the attribution of intelligence to a human is heavily context dependent. Sitting nearly naked on the ground and poking at ants with a stick would be regarded as odd in many contexts, but not when you're in the Kalahari Desert, looking for ant trails that might lead to a source of water. The world-class expert in any particular domain will not necessarily top the charts on a standard test of general intelligence.[8] Likewise, a computational device when taken out of its intended context of application can be useless—an extreme case being when it is dropped into a swimming pool. As a physical artifact, the device will still have affordances—it could plug a leak—but it would not interact with people on the basis of its computational capabilities. Or, try entering a chess machine in a natural language processing competition. Reductio ad absurdum examples such as these perhaps make the point less well than realistic examples. A geographic information system accepts data types x, y, and z, but the user needs (context of application) to integrate data types q, r, and s, which can't be done. Or, Box A cannot communicate with Box B. These sorts of problems are daily fare in complex sociotechnical workplaces plagued by mandated and legacy systems.

To continue with the triple analysis, replace the human who knows how to work with the computational device with a human who does not, and again the device becomes essentially useless. An extreme example would be a chess machine used by someone who knows nothing about games. The machine retains some of its affordances—the human could push buttons just to see what happens or could use the machine as a doorstop. To continue the triple, change the machine (its computational and interface capabilities), and you get entirely different patterns of interactions with the human and the context.

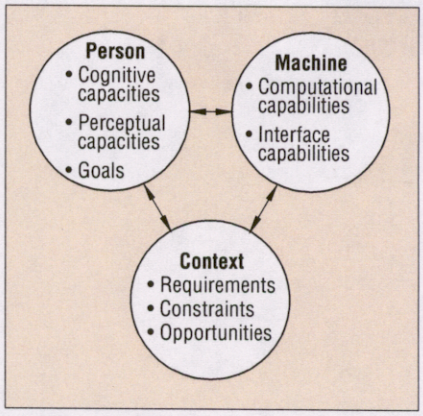


Figure 1. The triple.

Changes to the human or the context can render the machine useless or make the context inappropriate; the human is rendered less capable by changes to the context (for example, a PC user who finds himself surrounded by Macintosh users) or the machine. According to the Triples rule, intelligence emerges from the triple interaction. The rule thus runs head-on into the traditional definition of intelligence in AI: the Turing test.

Turing redux

Alan Turing's paper "Computing Machinery and Intelligence"[9] gave AI a vision and its first great challenge.[10] As Turing described it, the test is an imitation game that involves a man, a woman, and a judge, all communicating but unable to see one another. The judge's task is to decide which of the other two is the woman; the others try to persuade the judge that one of them is the woman and that the other is the man. Turing is usually understood to mean that the game should be played with the question of gender (being female) replaced by the question of species (being human). The judge is faced with the task of differentiating a human participant from a machine pretending to be human.

Borrowing from Patrick Hayes and Kenneth Ford,[11,12] we point out just three of the problems with the Turing test. First, from the standpoint of experimental design, it confounds the machine's intelligence with that of the judge. If a machine passes the test, is it demonstrated to be intelligent, or was the judge not intelligent enough to ask sufficiently telling questions? The game conditions say nothing about the judge, but the game's success depends crucially on

how intelligent the judge is.

Second, the Turing test confounds intelligence with cleverness. To pass the test, a machine would have to not only give a human-like impression but also be an expert on making a good impression. It would have to avoid exhibiting any inhuman talents that it might have; it would always have to lie, cheat, and dissemble. The winner of the Loebner competition, for example, sometimes deliberately mistyped a word and then backspaced to correct it at human typing speed. This strategy is clever, but such tricks should not be central to AI.

A third difficulty with the Turing test is that the definition of intelligence keeps shifting. As AI progresses and machines increasingly perform tasks previously considered to involve human intelligence, those abilities are no longer taken to be definitive. When Eliza first appeared, some people found its conversational abilities quite human-like. No machine until then could have reacted even in a simple way to what had been said to it. But during the Loebner competition, many programs were instantly revealed as nonhuman precisely by the first hint of their behavior's resemblance to Eliza's. The ability to perform simultaneous translation could soon be reduced to the merely mechanical. Turing tests have become circular: They define the qualities for which they are claiming to be evidence.

We could argue that the Turing test should not be regarded as the defining goal for AI but as a spur to technological progress—constantly pushing us to reach for Rene Descartes' dream of the "enlightened machine."[13] But why should we take it as our only goal to build something that is just like us? A dog would never win any imitation game, but there seems to be no doubt that dogs exhibit cognition, and a machine with the cognitive and communicative abilities of a dog would be an interesting challenge for AI.[14] More importantly, our most useful computer applications (including AI programs) are often valuable by virtue of their lack of humanity. There are cameras, copiers, televisions, automobiles, battery rechargers, and laptop operating systems all incorporating algorithms that use AI ideas and techniques but are not usually advertised as "intelligent" or "expert."[12]

If we abandon the Turing test as the defining goal for AI, the goal can shift from making artificial superhumans that can replace us to making artifacts that we can use to amplify and extend our cognitive abilities. AI should play a central role in this exciting new technology of cognitive prosthetics,[15] but to do so it must turn its back on the Turing test. What, then, should be the test?

The Triples rule suggests that the problem here is with the question itself—there is no single test. In fact, there are boundless numbers of tests. In general, for any task x or y that a human or machine can do (each under the appropriate set of contextual constraints), the human–machine–context

PURPOSE The IEEE Computer Society is the world's largest association of computing professionals, and is the leading provider of technical information in the field.

MEMBERSHIP Members receive the monthly magazine **COMPUTER**, discounts, and opportunities to serve (all activities are led by volunteer members). Membership is open to all IEEE members, affiliate society members, and others interested in the computer field.

BOARD OF GOVERNORS

Term Expiring 2002: Mark Grant, Gene F. Hoffnagle, Karl Reed, Kathleen M. Swigger, Ronald Waxman, Michael R. Williams, Akihiko Yamada

Term Expiring 2003: Fiorenza C. Albert-Howard, Manfred Broy, Alan Clements, Richard A. Kemmerer, Susan A. Mengel, James W. Moore, Christina M. Schober

Term Expiring 2004: Jean M. Bacon, Ricardo Baeza-Yates, Deborah M. Cooper, George V. Cybenko, Wolfgang K. Giloi, Haruhisha Ichikawa, Thomas W. Williams

Next Board Meeting: 10 May 02, Portland OR

IEEE OFFICERS

President: RAYMOND D. FINDLAY
President-Elect: MICHAEL S. ADLER
Past President: JOEL B. SYNDER
Executive Director: DANIEL J. SENESE
Secretary: HUGO M. FERNANDEZ VERSTAGEN
Treasurer: DALE C. CASTON
VP, Educational Activities: LYLE D. FEISEL
VP, Publications Activities: JAMES M. TIEN
VP, Regional Activities: W. CLEON ANDERSON
VP, Standards Association: BEN C. JOHNSON
VP, Technical Activities: MICHAEL R. LIGHTNER
President, IEEE-USA: LeEARL A. BRYANT

IEEE COMPUTER SOCIETY

EXECUTIVE COMMITTEE

President: WILLIS K. KING*
University of Houston
Dept. of Comp. Science
501 PGH
Houston, TX 77204-3010
Phone: +1 713 743 3349 Fax: +1 713 743 3335
w.king@computer.org

President-Elect: STEPHEN L. DIAMOND*
Past President: BENJAMIN W. WAH*
VP, Educational Activities: CARL K. CHANG *
VP, Conferences and Tutorials: GERALD L. ENGEL*
VP, Chapters Activities: JAMES H. CROSS
VP, Publications: RANGACHAR KASTURI
VP, Standards Activities: LOWELL G. JOHNSON (2ND VP)*
VP, Technical Activities: DEBORAH K. SCHERRER(1ST VP)*
Secretary: DEBORAH M. COOPER*
Treasurer: WOLFGANG K. GILOI*
2001–2002 IEEE Division VIII Director: THOMAS W. WILLIAMS
2002–2003 IEEE Division V Director: GUYLAINE M. POLLOCK
Executive Director: DAVID W. HENNAGE

*voting member of the Board of Governors

IEEE

COMPUTER SOCIETY WEB SITE

The IEEE Computer Society's Web site, at **http://computer.org**, offers information and samples from the society's publications and conferences, as well as a broad range of information about technical committees, standards, student activities, and more.

COMPUTER SOCIETY OFFICES

Headquarters Office
1730 Massachusetts Ave. NW
Washington, DC 20036-1992
Phone: +1 202 371 0101 • Fax: +1 202 728 9614
E-mail: hq.ofc@computer.org

Publications Office
10662 Los Vaqueros Cir., PO Box 3014
Los Alamitos, CA 90720-1314
Phone: +1 714 821 8380
E-mail: help@computer.org
Membership and Publication Orders:
Phone: +1 800 272 6657 Fax: +1 714 821 4641
E-mail: help@computer.org

European Office
13, Ave. de L'Aquilon
B-1200 Brussels, Belgium
Phone: +32 2 770 21 98 • Fax: +32 2 770 85 05
E-mail: euro.ofc@computer.org

Asia/Pacific Office
Watanabe Building
1-4-2 Minami-Aoyama, Minato-ku,
Tokyo 107-0062, Japan
Phone: +81 3 3408 3118 • Fax: +81 3 3408 3553
E-mail: tokyo.ofc@computer.org

EXECUTIVE STAFF

Executive Director: DAVID W. HENNAGE
Publisher: ANGELA BURGESS
Assistant Publisher: DICK PRICE
Director, Volunteer Services: ANNE MARIE KELLY
Chief Financial Officer: VIOLET S. DOAN
Director, Information Technology & Services: ROBERT CARE
Manager, Research & Planning: JOHN C. KEATON

triple is "intelligent" if the task can be conducted better than if either the human or the machine were to conduct the task without engaging in a partnership with the other.[16] The word "better" is deliberately left open in this formulation. It might be instantiated by "more efficiently," "more rapidly," or "more economically," but it might just as well mean "more playfully." This is how it should be if the machine is human-centered. There is an appropriate fit between the human, the machine, and the context such that the human's cognitive, perceptual, and collaborative capacities are enhanced. AI thus becomes Amplified Intelligence.

This line of thinking helps put to rest the doomsaying about intelligent machines taking over the world, which is another story.[14] This is just the first of what we hope will become a series of essays that discuss candidate principles for this thing called human-centered computing. The Triples rule is one of the first such principles. We invite others to suggest additional ones. ■

References

1. S. Wyatt, J.A. Fraser, and F.G.L. Stock, *The Effects of Monotony in Work*, Report no. 56, Industrial Fatigue Research Board, His Majesty's Stationery Office, London, 1929.
2. M.S. Viteles, *Industrial Psychology*, Norton, New York, 1932.
3. E.M. Newbold, *A Contribution to the Study of the Human Factor in the Causation of Accidents*, Report no. 34, Industrial Fatigue Research Board, His Majesty's Stationery Office, London, 1926.
4. H.G. Weston, "A Note on Machine Design in Relation to the Operative," *3rd Ann. Report Industrial Fatigue Research Board, 31 December 1922*, Industrial Fatigue Research Board, His Majesty's Stationery Office, London, 1923, pp. 71–75.
5. D.D. Woods, "Coping with Complexity: The Psychology of Human Behavior in Complex Systems," *Mental Models, Tasks, and Errors*, L.P. Goodstein, H.B. Anderson, and S.E. Olsen, eds., Taylor and Francis, London, 1988, pp. 128–148.
6. D.D. Woods and D. Tinapple, "Presidential Address: Watching Human Factors People at Work," *Proc. 43rd Ann. Meeting Human Factors and Ergonomics Soc.*, Human Factors and Ergonomics Soc., 1998; http://csel.eng.ohio-state.edu/hf99.
7. K. Vicente, *Cognitive Work Analysis*, Lawrence Erlbaum, Mahwah, N.J., 2000.
8. R.R. Hoffman, "How Can Expertise Be Defined? Implications of Research from Cognitive Psychology," *Exploring Expertise*, R. Williams, W. Faulkner, and J. Fleck, eds., Macmillan, New York, 1998, pp. 81–100.
9. A.M. Turing, "Computing Machinery and Intelligence," *Mind*, vol. 59, no. 236,1950, pp. 433–460.
10. A. Hodges, *Alan Turing: The Enigma*, Simon and Schuster, New York, 1983.
11. P. Hayes and K.M. Ford, "Turing Test Considered Harmful," *Proc. 14th Int'l Joint Conf. Artificial Intelligence*, Morgan Kaufman, San Francisco, 1995, pp. 972–977.
12. K.M. Ford and P.J. Hayes, "On Computational Wings: Rethinking the Goals of Artificial Intelligence, *Scientific Am. Presents*, vol. 9, no. 4, Winter 1998, pp. 78–83.
13. R.R. Hoffman, P. Hayes, and K.M. Ford, "Human-Centered Computing: Thinking In and Out of the Box," *IEEE Intelligent Systems*, vol. 16, no. 5, Sept./Oct. 2001, pp. 76–79.
14. P.A. Hancock, *Essays on the Future of Human-Machine Systems*, Banta Information Services Group, Eden Prairie, Minn., 1997.
15. K.M. Ford, C. Glymour, and P.J. Hayes, "Cognitive Prostheses," *AI Magazine*, vol. 18, no. 3, Fall 1997, p. 104.
16. P.A. Hancock and S.F. Scallen, "Allocating Functions in Human-Machine Systems," *Viewing Psychology as a Whole*, R.R. Hoffman, M.F. Sherrick, and J.S. Warm, eds., Am. Psychological Association, Washington, D.C., 1998, pp. 509–540.



Robert R. Hoffman is a research scientist at the University of West Florida's Institute for Human and Machine Cognition and is a faculty associate at the university's Department of Psychology. He specializes in expertise studies, including knowledge elicitation and the design of human-centered systems. He received his BA, MA, and PhD in experimental psychology from the University of Cincinnati. He is a Fulbright Scholar and a fellow of the American Psychological Society. Contact him at the Inst. for Human & Machine Cognition, 40 Alcaniz St., Pensacola, FL 32501; rhoffman@ai.uwf.edu.



Patrick Hayes is a senior research scientist and the John Pace, Jr., Eminent Scholar at the University of West Florida. His research interests include knowledge representation and automatic reasoning, especially the representation of space and time and diagrammatic representations, and the philosophical foundations of AI and computer science. He received his BA in mathematics from Cambridge University and his PhD in artificial intelligence from the University of Edinburgh. Contact him at the Inst. for Human & Machine Cognition, 40 Alcaniz St., Pensacola, FL 32501; phayes@ai.uwf.edu.



Kenneth M. Ford is the founder and director of the University of West Florida's Institute for Human and Machine Cognition. He has an interdisciplinary interest in understanding cognition in both humans and other machines. He is the editor in chief of the AAAI/MIT Press, executive editor of the *International Journal of Expert Systems*, and associate editor of the *Journal of Experimental and Theoretical Artificial Intelligence*, and is a *Behavioral and Brain Sciences* associate. He received his PhD in computer science from Tulane University. Contact him at the Inst. for Human & Machine Cognition, 40 Alcaniz St., Pensacola, FL 32501; kford@ai.uwf.edu.



Peter Hancock is a professor at Central Florida University and president of the Human Factors and Ergonomics Society. His research interests include transportation safety, air traffic control, workstation design, mental workload, intelligent interfaces, simulator design, and other topics. He received a PhD in motor performance from the University of Illinois. Contact him at Dept. of Psychology, Univ. of Central Florida, Orlando, FL 32816; phancock@pegasus.cc.ucf.edu.

EFFECTS OF INFORMATION AND DECISION AUTOMATION ON MULTI-TASK PERFORMANCE

Ericka Rovira, Marla Zinni, & Raja Parasuraman
The Catholic University of America

Automation purported to assist human operators may itself be an additional source of complexity and uncertainty. Because high reliability cannot always be assured, imperfect automation can add to uncertainty and thereby degrade performance. The present study examined the relative benefits and costs of information and decision automation and investigated the effects of uncertainty resulting from automation unreliability during multiple task performance. Subjects were either provided with status information ("information" automation) or a recommendation for action ("decision" automation) for the system monitoring sub-task of the Multi-Attribute Task Battery (MAT). Two levels of automation reliability were compared. The detrimental effect of unreliable automation—a decrease in the detection rate of malfunctions—was greater for automation of higher reliability, a result consistent with previous findings of automation-related complacency. This effect of automation unreliability was also greater for decision than for information automation.

INTRODUCTION

Information-management and decision support tools for human operators of complex systems have been developed for a number of domains (Parasuraman & Mouloua, 1996). Although these automation aids may enhance operator and system performance, new error forms and degraded performance have also been noted (Billings, 1997; Parasuraman & Riley, 1997; Sarter, Woods, & Billings, 1997; Wickens, Gempler, & Morphew, 2000; Wiener & Curry, 1980). Problems in human-automation interaction have included unbalanced workload, reduced system awareness, mistrust, over-reliance and complacency, decision biases, and cognitive skill reduction (Parasuraman & Riley, 1997).

The paradox is that automation purported to assist the human operator may itself be an additional source of complexity and uncertainty. Moreover, because high reliability cannot always be assured, especially in dynamic environments, imperfect automation can add to uncertainty and thereby degrade performance. At least two forms of uncertainty are important, *information* uncertainty, in which the appropriate information sources relevant to the problem are not known or present only incomplete data, and *decision* uncertainty, where the choice between action options may not be clear, or the available options may be incomplete.

The present study examined the benefits and costs of automation with respect to these two forms of uncertainty. Parasuraman, Sheridan, and Wickens (2000) recently outlined a model of automation and human performance which also broadly distinguishes between automation applied to information and decision functions. There is some evidence that the costs of unreliability may vary between automation support applied to different stages of processing. Crocoll and Coury (1990) found that compared to manual performance, automation that provided only status information on a target in a simulated air defense task ("information automation") was equally beneficial to that which gave a recommendation to engage a specific target ("decision automation"). But when the automation was unreliable, target detection performance was worse with decision than with information automation. Sarter and Schroeder (2001) obtained similar results with pilots using a decision support system for de-icing aircraft. It is not known, however, whether similar findings would occur in situations in which automation reliability varies and when the operator is engaged in multiple tasks. Under such conditions, operators are typically poor in monitoring the automated task for occasional malfunctions, an effect referred to as automation complacency (Parasuraman, Molloy, & Singh, 1993). The goal of the present study was to examine the relative benefits and costs of information and decision automation on multi-task performance and complacency, and to investigate the effects of uncertainty resulting from automation unreliability.

METHOD

Participants

Twenty young adults (10 male, 10 female) volunteered and were each paid $15/hour for their participation. The age range of participants was 18-24 years (M = 20.9 years, SD = 1.8 years).

Multi-Attribute Task (MAT) Battery

A modified version of the Multi-Attribute Task Battery (MAT) (Comstock and Arnegard, 1992) was run on an IBM compatible computer, using a 14-inch VGA color monitor, and an analog joystick. The version used in the present study consisted of three main tasks: (1) engine systems monitoring; (2) two-dimensional compensatory tracking; and (3) fuel management. The tracking and fuel management tasks were always carried out manually and were designed to serve as loading tasks on the participant. The engine systems monitoring task was carried out either manually or under different forms of automation support.

The system monitoring task consisted of four vertical gauges with moving pointers, and green and red warning lights. The scales for the gauges were marked indicating the temperature and pressure for two aircraft engines. Occasional system malfunctions occurred at random intervals, indicated by the pointer on one of the four engine gauges deviating beyond acceptable limits. There were 16 such malfunctions during the 10 minute block of the manual condition and 32 malfunctions during the other 20 minute blocks. In all conditions, participants were required to correct a gauge when the pointers moved out of range by designating the correct gauge for correction as well as the direction for correction.

A two-dimensional compensatory tracking task with joystick control was presented in one window of the display. Dashed x- and y- axes were provided for reference. A green circular target symbol representing the deviation of the aircraft from its course fluctuated within the window in the x- and y- directions. The participant's task was to keep the aircraft within the central rectangle by applying the appropriate control inputs in the x- and y- directions.

The fuel management task was a simulation of the actions needed to manage the fuel system of the aircraft. The display for this task consisted of six rectangular regions that represented aircraft fuel tanks. Along the 'fuel lines' that interconnected these tanks were fuel pumps capable of transferring fuel in one direction, and at a specified flow rate. Participants were required to maintain a specific fuel level within both of the main tanks, by selectively activating pumps to keep pace with the fuel consumption in these main tanks.

Automation Support

Automation of two types was provided to assist the user in performing the system-monitoring task. During information automation, the automation assisted the participant by indicating which of the four gauges contained a possible malfunction, but did not indicate the corrective action. Under decision automation, the automation indicated not only which of the gauges contained a malfunction but also the direction of correction, up or down, that had to be carried out for a particular gauge. Additionally, automation reliability was manipulated across blocks across 2 levels, low (50% correct) and high (75% correct).

Procedure

Participants were given 40 minutes of instruction and training prior to the beginning of data collection. Each individual received verbal instruction on each task of the MAT, followed by a five-minute practice session for each task separately. They then proceeded to practice performing all three tasks simultaneously, while being instructed to give equal attention to each.

The main experimental session began with a 10-minute manual scenario in which participants were instructed to perform all three tasks manually, giving equal attention to each task. Participants were then introduced to the automation support. They were instructed that the automation was designed to help them with the system-monitoring task so that they could perform the primary tasks of tracking and fuel management. They were informed that the automation was not 100% reliable in all conditions. Each participant performed a five-minute session in which the automation was 100% reliable followed by two 20-minute (75% and 50% reliability) sessions of either information or decision automation and another 100% reliable block followed by two scenarios of the other type of automation. The order of type of automation (information or decision) and automation reliability (75% or 50%) was counterbalanced.

RESULTS

Engine Systems Monitoring Task

The mean rate of correct detections of engine malfunctions was computed for each participant under manual performance and for the different automation conditions. Under automation support (75% and 50% reliability only), detection rates were computed separately for reliable and unreliable trials within each automation support block. Similarly, the mean of the median reaction times (RTs) for correct detection were computed for manual performance, reliable automation, and unreliable automation.

We first analyzed the overall detection rates and RTs for the manual, reliable automation, and unreliable automation conditions. Analysis of variance of the detection rates with conditions as a repeated-measures factor showed that the effect of condition was significant, $F(2,38) = 33.7$, $p < .0001$. Application of the Newman-Keuls test showed that each condition differed significantly from the other ($p < .05$). Participants detected 87.1% of system malfunctions under manual performance while simultaneously performing the tracking and fuel management tasks. Under reliable automation support, there was a significant performance benefit of about 10%. When the automation was unreliable, however, detection performance was reduced to 63.6%. Note that the performance degradation was such that detection was not returned to the manual performance level but was further reduced, corresponding to a cost of about 27%.

Very similar results were obtained for detection RT. The main effect of condition was significant, $F(2,38) = 20.4$, $p < .0001$. Application of the Newman-Keuls test showed that all three means were significantly different from one another ($p < .05$). Compared to manual performance, reliable automation resulted in about a 27% speed up in detection time. When the automation was unreliable, however, detection speed was slowed by about 20%.

We next examined detection performance for reliable and unreliable automation trials separately as a function of the level of automation reliability (low or high) and the type of automation support (information or decision automation). Mean values for reliable trials are shown in Figure 1A/B. For reliable automation, the system malfunction detection rate was very good (> 94.5%) and did not vary significantly with either the level of reliability or automation type ($F(1,19) < 1$ in each case). Automation reliability also had no effect on detection RT ($F(1,19) < 1$). However, RT was significantly reduced for decision automation compared to information automation ($F(1,19) = 8.8$, $p < .01$).

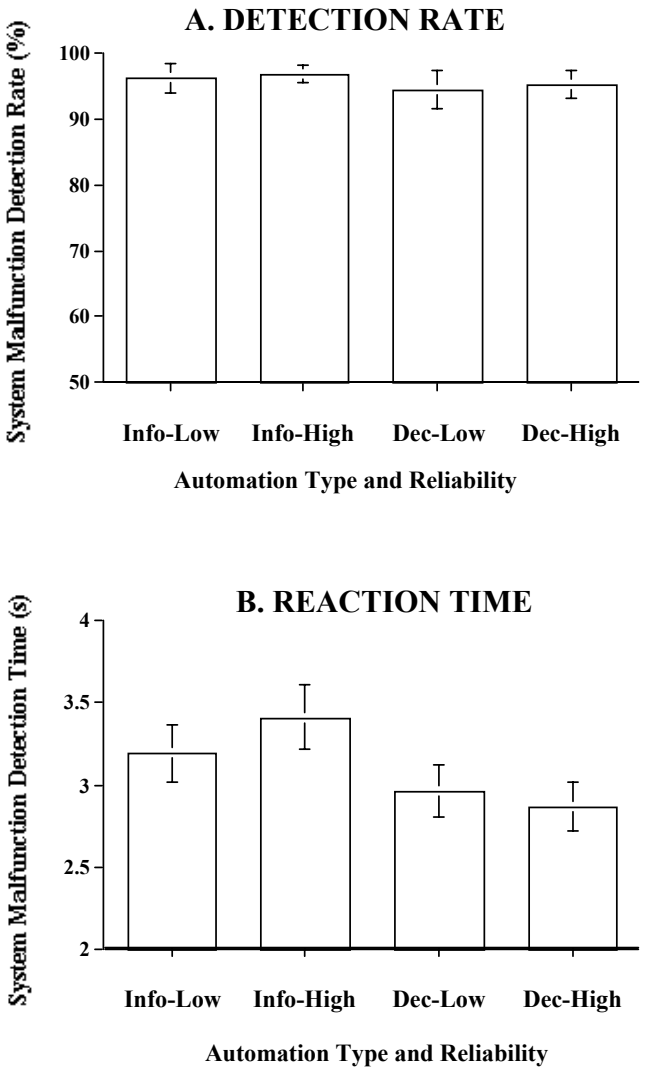


Figure 1: Reliable Automation

For the unreliable trials, the effect of reliability level was significant ($F(1,19) = 8.3$, $p < .01$). Whereas there was no significant effect of automation type ($F(1,19) = 1.6$, p = 22), there was a trend for a reliability X automation type interaction ($F(1,19) = 3.1$, $p = .09$). As Figure 2A shows, detection rates were equivalent for the low and high reliability conditions for information automation and for the low reliability decision automation condition. Detection rate under high reliability decision automation, however, was lower than in the other three conditions. Finally, analysis of the RTs for unreliable trials revealed no significant effects (see Figure 2B).

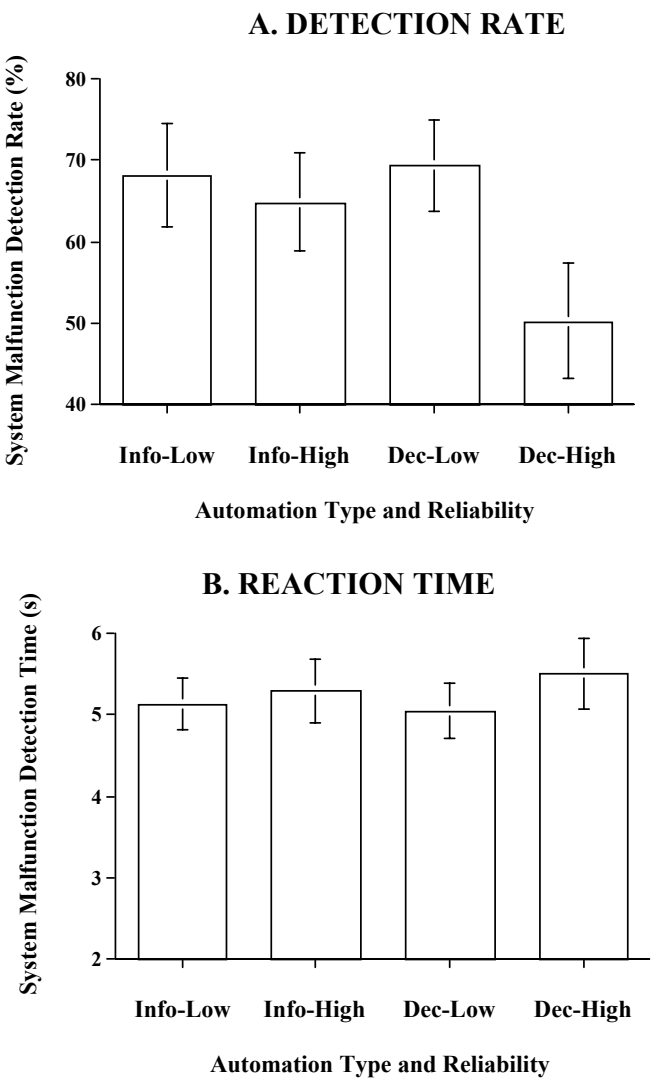


Figure 2: Unreliable Automation

Tracking and Fuel Management Task Performance

Performances on the two loading tasks were quantified to examine first whether the automation manipulation was sensitive to multi-task performance, and second, whether there were any performance tradeoffs with variations in the automation conditions. When all three tasks were performed manually, the mean tracking RMS error was 273.9 (*SE* = 15.0). When the engine monitoring task was automated, the mean tracking RMS error was 211.1 (*SE* = 13.8), a significant reduction in error ($t(19) = 5.5$, $p < .0001$). Mean fuel RMS scores for the manual and automation conditions were 136.0 (*SE* = 23.4) and 88.1 (*SE* = 12.0), respectively. These values were significantly different from one another ($t(19) = 2.2$, $p < .05$). Thus performance on both the tracking and fuel tasks were sensitive to the automation of the engine monitoring task, but there were no significant effects on performance of these tasks of automation type or reliability.

DISCUSSION

The present study examined the relative benefits and costs of information and decision automation and investigated the effects of uncertainty resulting from automation unreliability during multiple task performance. The results showed that reliable automation led to a significant benefit over manual performance. However, when the automation was unreliable, performance was lower than that of the manual condition, resulting in a performance cost of automation. Further, with respect to the distinction between information and decision automation, when automation was reliable no differences were found in detection rates across the groups. A benefit in reaction time emerged under decision automation in both low and high reliability conditions. Under trials in which automation was unreliable, performance was worse than that of reliable trials, with a greater drop in detection performance for decision than for information automation. This finding suggests an over-reliance on the automation when decision-making was automated. These results are consistent with those of previous findings (Crocoll & Coury, 1990; Sarter and Schroeder, 2001) but extend them to multi-task conditions in which operators have to distribute their attention between many different manual and automated tasks. Furthermore, the results show that the detrimental effects of unreliable automation are greater for automation of higher reliability, a finding consistent with previous findings of automation-related complacency (Parasuraman et al., 1993).

The paradox of imperfect automation is that automation that is often but not *always* accurate has a greater detrimental effect on operator decision making than when it is less frequently correct. Since perfect reliability cannot be reached, these findings need to be taken into consideration not only in design but also in training to combat these effects. To reduce the effects of complacency, control could be returned to the operator for a period of time to prevent de-skilling and encourage more attentional sampling of the automated task. Future research should evaluate methods to mitigate complacency especially under conditions of high levels of automation implementation.

ACKNOWLEDGMENTS

This work was supported in part by Grant DAAD17-00-p-0366 from the Army Research Laboratory, Ft. Huachuaca and in part by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

REFERENCES

Billings, C.E. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Comstock, J.R. & Arnegard, R.J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research.* Technical Memorandum 104174. NASA Langley Research Center, Hampton, VA.

Crocoll, W. M., & Coury, B., G. (1990). Status or recommendation: Selecting the type of information for decision aiding. *Proceedings of the 34th Annual Meeting of the Human Factors and Ergonomics Society (pp. 1524-1528).* Santa Monica, CA: Human Factors Society.

Parasuraman, R., Molloy, R., & Singh, I. (1993) Performance consequences of automation-induced "complacency". *The International Journal of Aviation Psychology, 3, 1-13.*

Parasuraman, R. & Mouloua, M. (1996) (Eds.). *Automation and human performance: Theory and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, & Cybernetics, 30*, 286-297.

Sarter, N., Woods, D. D. & Billings, C.E. (1997). Automation surprises in Handbook of Human Factors and Ergonomics, 2nd ed., G. Salvendy, Ed. New York: Wiley, pp. 1926-1943.

Sarter, N. & Schroeder, B. K. (2001). Supporting decision-making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors*, *43(4)*, 573-583.

Wickens, C. D., Gempler, M., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors, 2,* 99-126.

Wiener, E. L. & Curry, R. E. (1980). Flight-deck automation: promises and problems. *Ergonomics, 23,* 995-1011.

EFFECTS OF UNRELIABLE AUTOMATION ON DECISION MAKING IN COMMAND AND CONTROL

Ericka Rovira, Kathleen McGarry, & Raja Parasuraman

The Catholic University of America

The effectiveness of automated decision aids used by human operators in command and control systems may depend not only on automation reliability, but also on the type (stage) and level the automated support provides. Automation can be applied to information acquisition, information integration and analysis, decision choice selection, or action implementation (Parasuraman, Sheridan, & Wickens, 2000). The present study examined the effects of variations in the stage of automation support on performance in a "Sensor to Shooter" targeting simulation of command and control. Independent variables included the type and level of automation support (complete listing, priority listing, top choices, and recommendation of decision choice) and the reliability of the automation (60% and 80%). Dependent variables included accuracy and reaction time of target engagement decisions. Compared to manual performance, reliable automation did not affect the accuracy of target engagement decisions but did significantly reduce decision times. When the automation was unreliable, under the higher reliability condition (80%) there was a greater cost in accuracy performance for higher levels of automation aiding (priority listing, top choice, and recommendation) than at a lower level (complete listing). The results support the view that automation unreliability has a greater performance cost for decision automation than for information automation. This performance cost generalizes across a number of different forms of decision-aiding.

INTRODUCTION

The complexity of modern military systems is such that system efficiency can be compromised without the use of automation to support the human operators of these systems. This is because human operators of such systems are faced with making complex decisions under time stress and often with incomplete information. Real-time decision aids have been developed for a number of problem domains in both civilian and military applications. Unfortunately, automated aids have not always enhanced system performance, primarily due to problems in their use by human operators or to unanticipated interactions with other sub-systems. Problems in human-automation interaction have included unbalanced workload, reduced system awareness, decision biases, mistrust, over-reliance, and complacency (Parasuraman & Riley, 1997). There is thus a pressing need for designing automation that supports military operators in command and control in ways that avoid such negative influences.

It has been proposed that a taxonomy of automation can be applied to a four-stage model of human information processing: 1) information acquisition; 2) information analysis; 3) decision and action selection; 4) action implementation (Parasuraman et al., 2000; Wickens, 2000). Previous research has examined the effects of some of these stages of automation on operator performance. Stage 2 involves cognitive functions including the integration of incoming data and inferential processes. Automation can assist the user in inference and recommend possible courses of action (Jones et al., 2000). Stage 3 pertains to selecting from various decision options. Automation of this stage is generally based on an inference of the state of the world (stage 2); however, different outcomes may be given certain values, hence making automation of this stage distinct from stage 2 (Wickens, 2000). Crocoll and Coury (1990) showed that stage 2 and stage 3 automation in an air defense task reduced the amount of time required to identify an aircraft. However, automation unreliability had a greater detrimental effect on performance when it provided an erroneous recommended action (stage 3) than when it gave only the wrong status information (stage 2). In a study on decision-aiding in the context of in-flight icing, Sarter and Schroeder (2001) also showed that users given unreliable status information (stage 2) were slower to respond, but made fewer errors than users given stage 3 automation.

The present study examined the effects of different stages of automation and automation reliability on user performance in a simulation of a Sensor-to-Shooter system (STS), which evaluates some of the tasks involved in military command and control. The effects of automation at stages 2 and 3 in the automation taxonomy, both which are typically characterized under the heading of decision-aiding (Wickens, 2000), were examined. Two factors differentiate the present study from the previous studies by Crocoll and Coury (1990) and Sarter and Schroeder (2001). First, the effects of two levels of automation reliability were examined. Decision-aiding automation in command and control systems can be imperfect for a number of reasons, including insensitivity to context. This allowed an assessment of the range of automation unreliability effects on different aspects of decision making, as well as a comparison to previous

work on automation complacency (Parasuraman, Molloy, and Singh, 1993) in which operator reliance on imperfect automation is evaluated. Second, because unreliability of stage 3 automation has previously been shown to have performance costs, we examined three different types of automation support within this stage of automation, to examine the generality of the phenomenon.

METHODS

Participants

Eighteen undergraduate students aged 18 to 22 years (M=19.83, S.D. = 1.25) from the Catholic University of America volunteered and were paid $15 per hour for their participation.

Apparatus

A PC-based medium-fidelity software simulation of a sensor to shooter targeting system (STS) was used. The program runs on a Pentium-II processor with a 17-inch monitor attached. A mouse was used as the input device.

Task and Procedures

The participants were required to identify the most dangerous enemy target and to select a corresponding friendly unit to engage in combat with the target. An action selection was required within 10 seconds. A 4 (automation levels) x 2 (automation reliability) within-subjects design was used. The four levels of automation included: complete listing, which corresponds to stage 2 automation, and three different levels within stage 3 automation, priority listing, top choices, and recommendation automation. The complete listing automation provided the participant with a complete list of all possible engagement combinations including the distances between enemy targets, friendly units, and headquarters. Because no explicit pointer to decision choice was provided, this corresponds to information (stage 2) automation. The priority listing automation provided the participant with a complete list of all possible engagement combinations including the distances between enemy targets, friendly units, and headquarters; however, in this instance the listings were prioritized with the best selection first and the worst choice last. The top choices automation provided the best three choices for engagement, including the distances between enemy targets, friendly units, and headquarters. The recommendation automation suggested the best enemy-friendly engagement; however, no distance information was provided. Figure 1 shows the Sensor to Shooter interface without automation support.

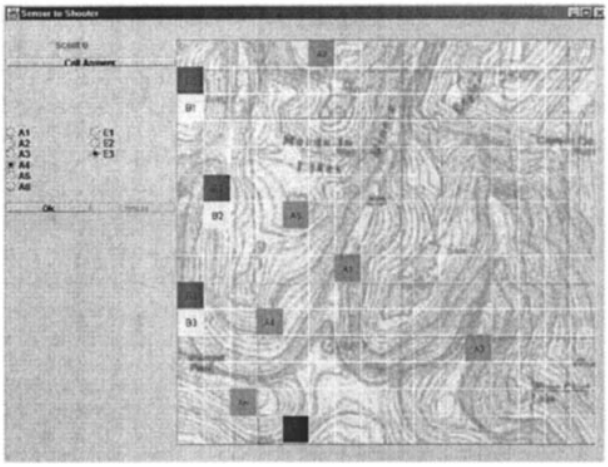


Figure 1: Sensor To Shooter Interface (Manual)

Automation reliability was varied across two values (60% and 80%) by manipulating the proportion of trials in which a correct assessment or decision was provided. In the 60% automation reliability condition, 24 trials were reliable (correct automated assessment) and 16 trials were unreliable (incorrect automated assessment). Similarly, in the 80% reliability condition 32 trials were reliable and eight trials were unreliable. The reliability was manipulated for the different levels of decision-aiding. Participants were informed that while the automation was highly reliable, it was not 100% reliable all of the time.

Each participant first performed the task manually (20 trials) with no automation support. All participants were presented with 40 static trials of each condition, accounting for ordering effects. However, participants completed both reliability conditions (60% and 80%) for each type of decision-aiding (complete listing, priority listing, top choices, and recommendation) before a new type of decision support was introduced. Additionally, 20 trials of 100% reliable automation were presented in between each condition, in an effort to recalibrate trust in the automated decision aid. In all, each participant completed 500 trials.

Dependent variables included accuracy and speed of enemy-friendly engagement selections. To obtain subjective measures of mental workload, participants were asked to fill out the NASA-TLX subjective scale. Scores were obtained by averaging the rating of the six TLX subscales, a method found to be psychometrically equivalent (Nygren, 1991) to the weighted subscore averaging suggested by the authors of the NASA-TLX. Subjective ratings of trust in the detection aid were obtained on a Likert-rating scale ranging from 0-100 built after scales used by Lee and Moray (1992, 1994).

RESULTS

Manual vs. Automation Performance

The mean rate of correct enemy-friendly engagement selections (decision accuracy) was computed for each participant under manual performance and for the

different automation conditions. Under decision-aiding automation (60% and 80% reliability), engagement selection rates were computed separately for reliable and unreliable trials within each automation condition. A paired samples *t*-test of engagement selection rates showed that accuracy of target engagement decisions did not differ significantly from manual performance when the users were provided with a reliable automation aid, $t(17) = .623, p > .05$. Accuracy for both manual and reliable automation was over 87%. However, there was a significant difference in accuracy between manual and unreliable decision-aiding, $t(17) = 6.9, p < .0001$; accuracy declined under unreliable automation. Figure 2 shows that there was no difference in accuracy between manual and reliable automation, and that under unreliable automation accuracy declined to 70%.

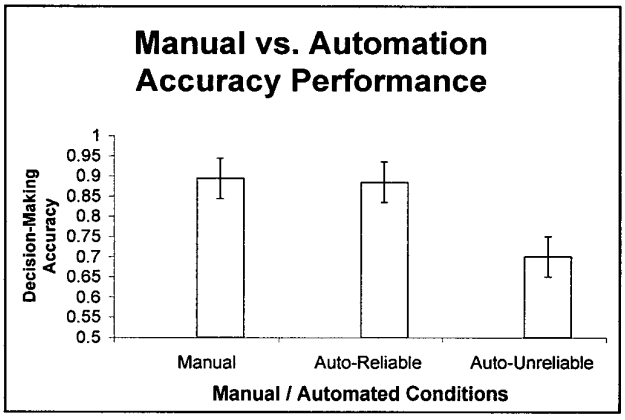


Figure 2: Manual vs. Automation Accuracy Performance

Enemy-friendly engagement reaction times (RTs) were reduced significantly from the manual condition when participants were given a reliable decision support tool, $t(17) = 5.3, p < .01$. Under the highest level of decision-aiding, recommendation automation, participants only had to click "okay" if they agreed with the automation. This may have led to a difference in RTs simply due to the design of the interface. Therefore a paired samples *t*-test of engagement selections between manual and reliable decision-aiding excluding the trials with the highest decision-aiding automation revealed that a reduction in RTs still occurred due to reliable automation, $t(17) = 2.37, p < .05$. Although RTs increased under unreliable decision-aiding, this increase was not significantly different from manual RTs, $t(17) = -.486, p > .05$. Figure 3 shows that compared to baseline, there was a significant decrease in RT with the use of a reliable aid and an increase in RT with the use of an unreliable aid.

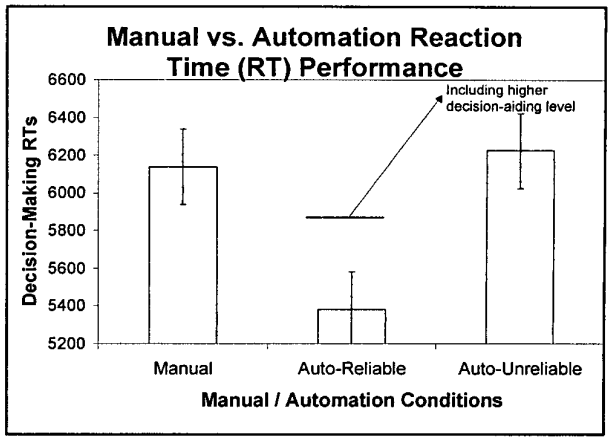


Figure 3: Manual vs. Automation Reaction Time (RT) Performance

Effects of Automation Level and Reliability

Reliable Trials. We next examined decision accuracy for reliable and unreliable decision-aiding trials separately as a function of the stage and level of automation (complete listing, priority listing, top choices, and recommendation) and automation reliability (60% and 80%). For the reliable trials, the effect of automation level was significant, ($F(3, 51) = 6.847, p < .01$), but that of automation reliability was not ($F(1, 17) = .054, p > .05$). The interaction between the two was significant, $F(3,51) = 2.9, p < .05$. Figure 4 shows that performance generally increased with the level of automation, being greatest for the recommendation condition and lowest for the complete listing condition. However, this increase was somewhat greater for the low (60%) than for the high (80%) reliability condition. In general, there was only moderate variability in performance across automation stages when automation was reliable.

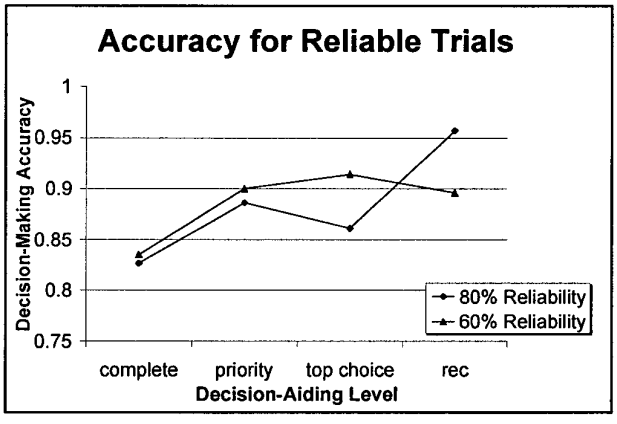


Figure 4: Accuracy for Reliable Trials

Unreliable Trials. For the unreliable trials, the effect of automation level ($F(3, 51) = 1.109, p > .05$) and automation reliability ($F(1, 17) = .939, p > .05$) were not significant, but the interaction between the two was, $F(3,51) = 14.392$, $p<.0001$. Figure 5 shows worse performance at 80%

automation reliability for the higher levels of decision-aiding (priority listing, top choice, and recommendation), whereas the opposite was true for complete listing decision-aiding. As indicated before, accuracy under manual and reliable automation exceeded 87%.

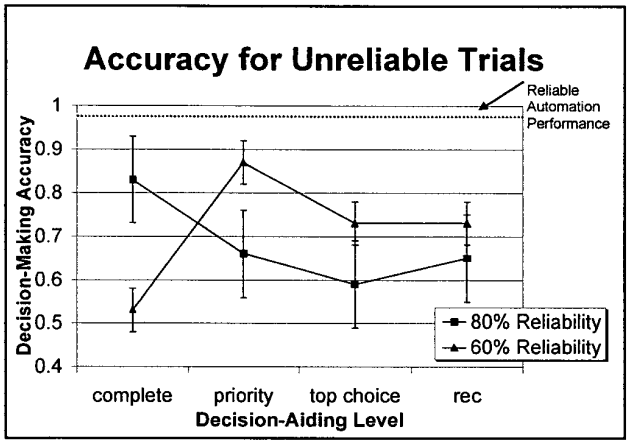


Figure 5: Accuracy for Unreliable Trials

Subjective Ratings of Mental Workload

A paired samples *t*-test revealed no significant difference in ratings of mental workload between manual (M = 31.11, SE = 4.15) and automated conditions (M = 34.12, SE = 4.13), $t(17) = -.89$, $p > .05$. There were significant effects of automation level ($F(3,51) = 4.66$, $p < .05$) and automation reliability ($F(1,17) = 7.89$, $p < .05$) on subjective workload ratings. The interaction between automation level and automation reliability was not significant, $F(3,51) = 1.64$, $p > .05$. Mental workload ratings were higher in the 80% automation reliability condition for complete listing, top choice, and recommendation decision-aiding, but not for priority listing. Collapsed across both reliability conditions, subjective ratings of mental workload were similar for complete listing, top choice, and recommendation decision-aiding, but were lower for priority listing,

Subjective Ratings of Trust and Self-Confidence

There was a significant effect of automation level on ratings of trust ($F(3,51) = 5.22$, $p < .05$); however, unexpectedly automation reliability had no significant effects on trust ($F(1,17) = .14$, $p > .05$). The interaction between automation level and automation reliability was not significant, $F(3,51) = 1.51$, $p > .05$. Collapsing across both reliability conditions, trust ratings were greatest for priority listing, followed by recommendation, then top choice, and lowest for complete listing decision-aiding. Automation level ($F(3,51) = .507$, $p > .05$) and automation reliability ($F(1,17) = .007$, $p > .05$) had no significant effects on ratings of self confidence. The interaction between automation level and automation reliability was also not significant, $F(3,51) = 1.72$, $p > .05$.

DISCUSSION

Operators of complex military systems are tasked with making multifaceted decisions under high time stress and often with uncertain information sources. Automation of decision making in command and control systems may be imperfect for a variety of reasons including enemy deception or faulty sensor readings.

The purpose of this study was to investigate the effects of unreliable decision-aiding on human performance in a command and control environment. As expected, automation unreliability did have an impact on decision-making performance. The view that unreliable automation has greater costs at stage 3 than stage 2 in the Parasuraman et al. (2000) automation taxonomy was verified (Crocoll & Coury, 1990; Sarter & Schroeder, 2001; Wickens, 2000) and generalized to different levels of automation within stage 3. Specifically, the results showed that under the high reliability condition (80%) there was a greater cost in accuracy performance at stage 3 (priority listing, top choice, and recommendation automation) than at stage 2 (complete listing automation).

The finding that unreliable automation has greater costs at stage 3 than stage 2 is potentially related to the kind of support the different automation types provide a user. As previously stated, stage 2 automation aids the user in integrating information, the inferential process, and may also recommend possible courses of action but does not explicitly support any particular decision choice or course of action, unlike stage 3 automation. Therefore, it is possible that under unreliable automation, performance is better at stage 2 because the user continues to generate the values for the different courses of action and hence is not as detrimentally influenced by an inaccurate decision aid. In contrast, users of stage 3 automation may no longer routinely develop or explore novel alternatives apart from those provided by the automation, as a result of which they may be penalized when the automation is inaccurate.

ACKNOWLEDGEMENTS

This work was supported in part by Grant DAAD17-00-p-0366 from Army Research Laboratory, Ft. Huachuaca and in part by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

REFERENCES

Crocoll, W. M., & Coury, B., G. (1990). Status or recommendation: Selecting the type of information for decision-aiding. *Proceedings of the 34th Annual Meeting of the Human Factors and Ergonomics Society (pp. 1524-1528).* Santa Monica, CA: Human Factors Society.

Jones, P. M., Wilkins, D.C., Bargar, R., Sniezek, J., Asaro, P., Danner, N., Eychaner, J., Chernyshenko, S., Schrah, G., Hayes, C., Tu, N., Ergan, H., & Lu, L. (2000). CoRAVEN: Knowledge-based support for intelligent analysis. In M. E. Benedict (Ed.), *ARL Federated Laboratory 4th Annual Symposium: Advanced Displays & Interactive Displays Consortium (*pp. 89-94).

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human machine systems. *Ergonomics, 35*, 1243-1270.

Lee, J., & Moray, N., (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human Computer Studies, 40, 153-184.*

Nygren, T.E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33, 17-31.*

Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance consequences of automation-induced "complacency". *The International Journal of Aviation Psychology, 3, 1-13.*

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, & Cybernetics, 30, 286-297*

Sarter, N., & Schroeder, B., K. (2001) Supporting decision-making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors, 43(4), 573-583*

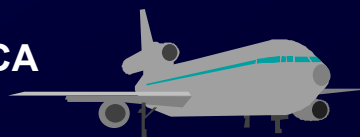
Wickens, C., D. (2000). Imperfect and unreliable automation and its implications for attention allocation, information access and situation awareness. Technical Report ARL-00-10/NASA-00-2

Adaptive Automation Matched to Human Mental Workload

Raja Parasuraman
Cognitive Science Laboratory
The Catholic University of America

NATO Advanced Research Workshop on Operator Functional Assessment and Impaired Performance in Complex Work, Il Ciocco, Italy, April 3-7, 2002

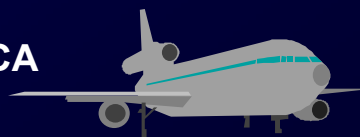




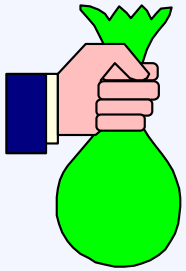
Overview

- Human Interaction with Automation
- Adaptive Automation (or Dynamic Function Allocation)
- Adaptation to Operator Mental Workload
- Cerebral Blood Flow Based Adaptive Automation
- Conclusions on Adaptive Systems

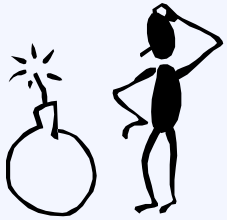




Reasons for the March Towards More Automation



- Cost



- Safety?
- Technical Capability



X
- Human Factors?

Systems

- Aviation and air traffic control
- Ground and maritime transportation
- Process control and manufacturing
- Military command and control
- Medicine and health care
- Drug design/Molecular genetics





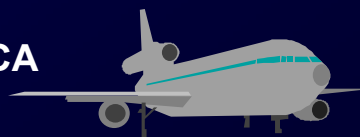
Cockpit Automation: Boeing 777



Automation and Human Performance: Benefits

- Improved precision of performance
- Operational flexibility
- Reduced mental workload
- Enhanced safety (automated warning systems)

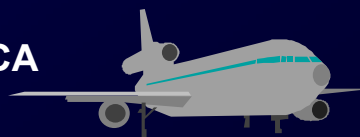




Some Potential Costs of Automation

- Loss of situation awareness
- Mode error/confusion
- Manual skill degradation
- Over-reliance and complacency
- Unbalanced workload





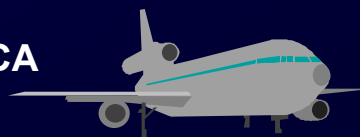
"Clumsy" Automation (Wiener, 1988)

Automation that does not reduce workload when operator most needs it; conversely, provides "aid" when none is needed.

Unbalanced Workload

- workload increased during high task load
- workload decreased during low task load



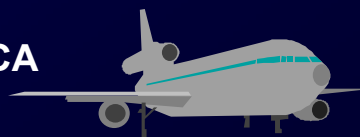


Adaptive Automation

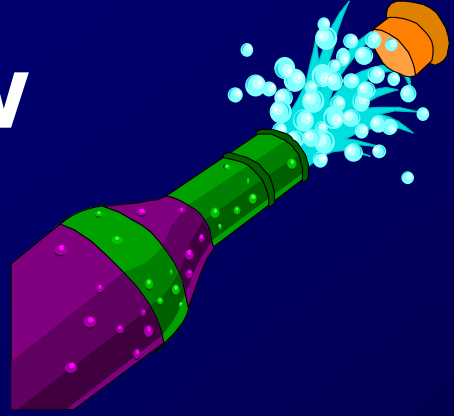
Human-computer systems in which the "division of labor"and/or the interface between human and machine agents is not fixed at system design but can vary dynamically during system operations

Dynamic Function Allocation





Adaptive Automation: New Wine in New Bottles



- Rouse (1976)—Dynamic Task Allocation
- Hancock et al. (1985)—Dynamic Task Allocation Model
- Lockheed—Pilot's Associate Program (1990)
- Parasuraman et al. (1992)–Adaptive Automation Research

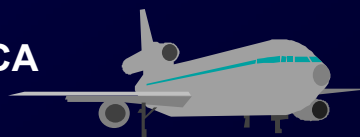
  Navy Adaptive Function Allocation for Intelligent Cockpits (AFAIC) Program

- Army—Rotorcraft Pilot's Associate Program (1998)
- 1992-Present: Empirical Studies (Hancock, Kaber, Miller, Parasuraman, Scerbo, …..etc)

Potential Benefits of Adaptive Automation

The costs sometimes associated with high-level automation—unbalanced workload, reduced situation awareness, complacency, skill degradation, etc.—may be mitigated by adaptive automation



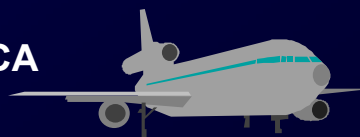


Varieties of Adaptive Automation

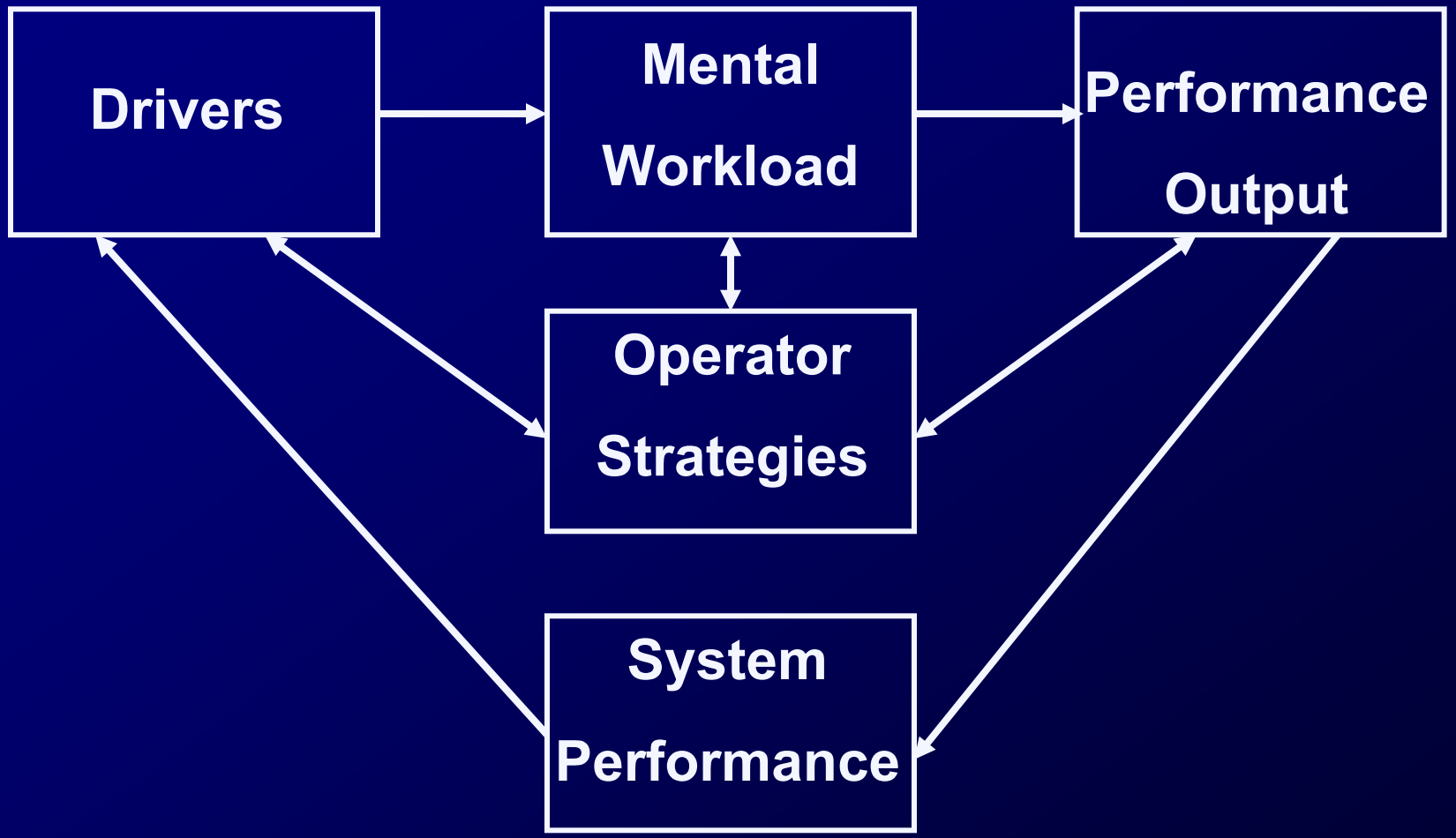
- Adaptive aiding of the human operator 
- Dynamic task allocation between human operator and computer systems 
- Context-dependent information management

Workload-Matched Adaptive Automation

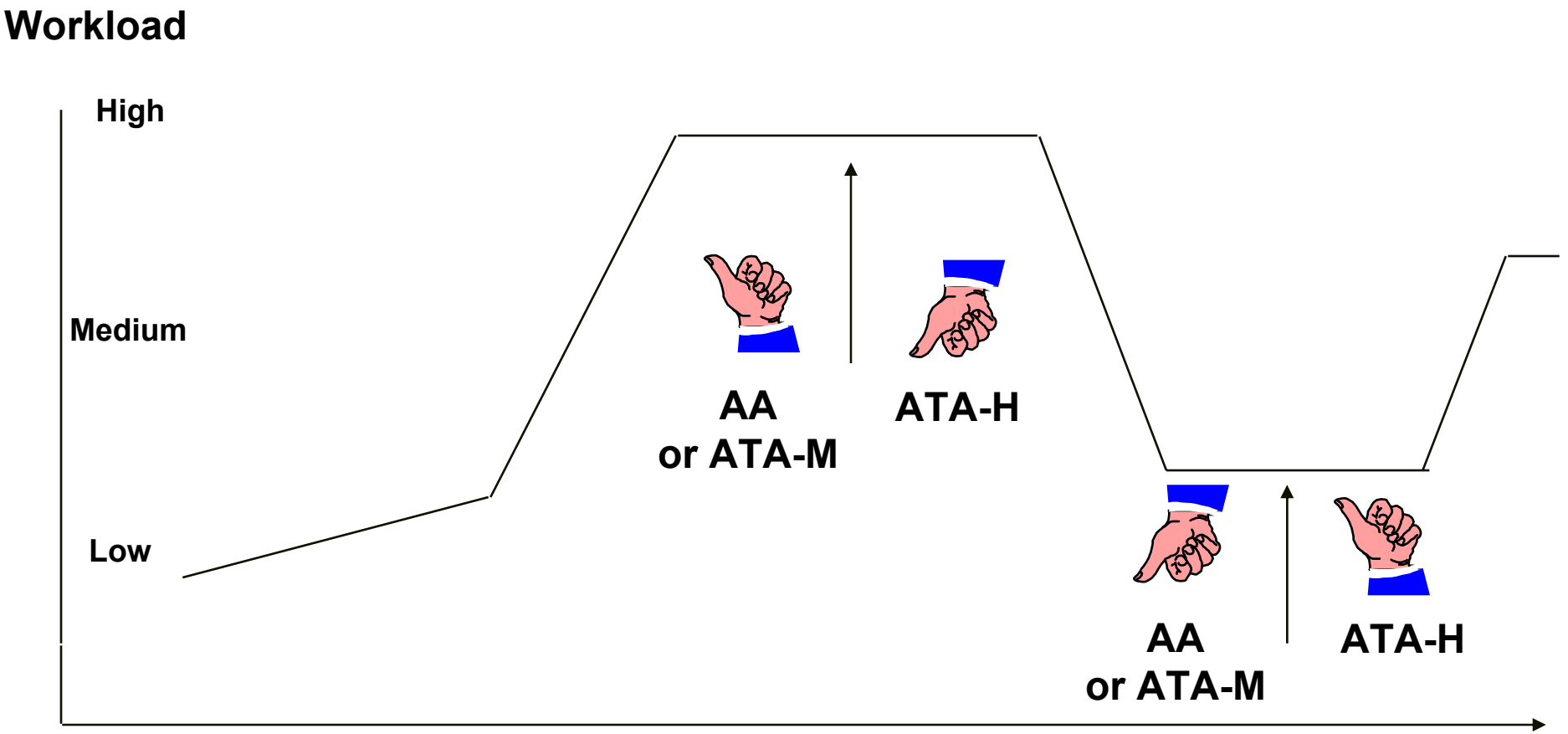




The Mental Workload Cycle



Workload-Matched Adaptive Automation



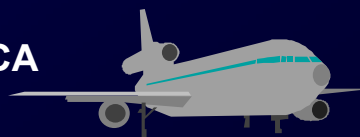
AA = Adaptive Aiding
ATA-H = Adaptive Task Allocation to Human
ATA-M = Adaptive Task Allocation to Machine

Optimal Time for Adaptation

Poor Time for Adaptation

Benefits of Adaptive Aiding

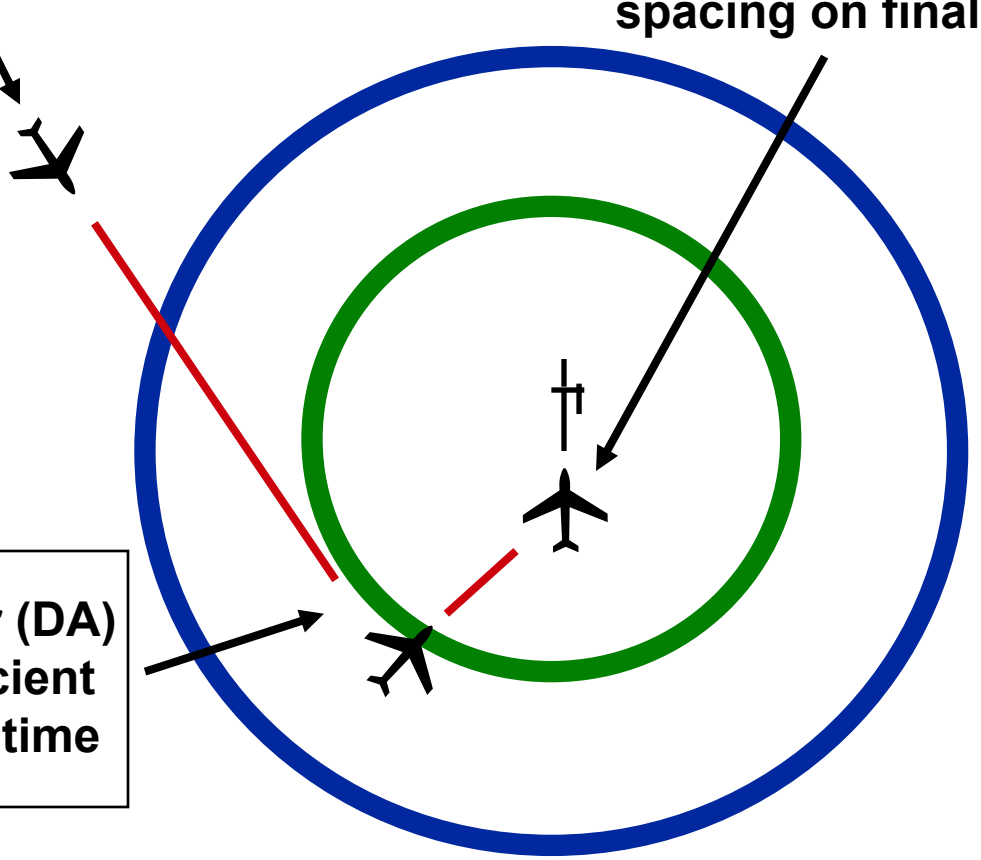




CENTER TRACON AUTOMATION SYSTEM (CTAS)—AUTOMATION FOR AIR TRAFFIC CONTROLLERS

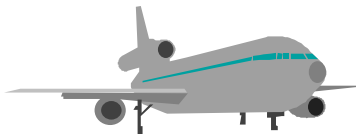
Traffic Management Advisor (TMA) plans sequence and landing time

Final Approach Spacing Tool (FAST) advises for accurate spacing on final approach

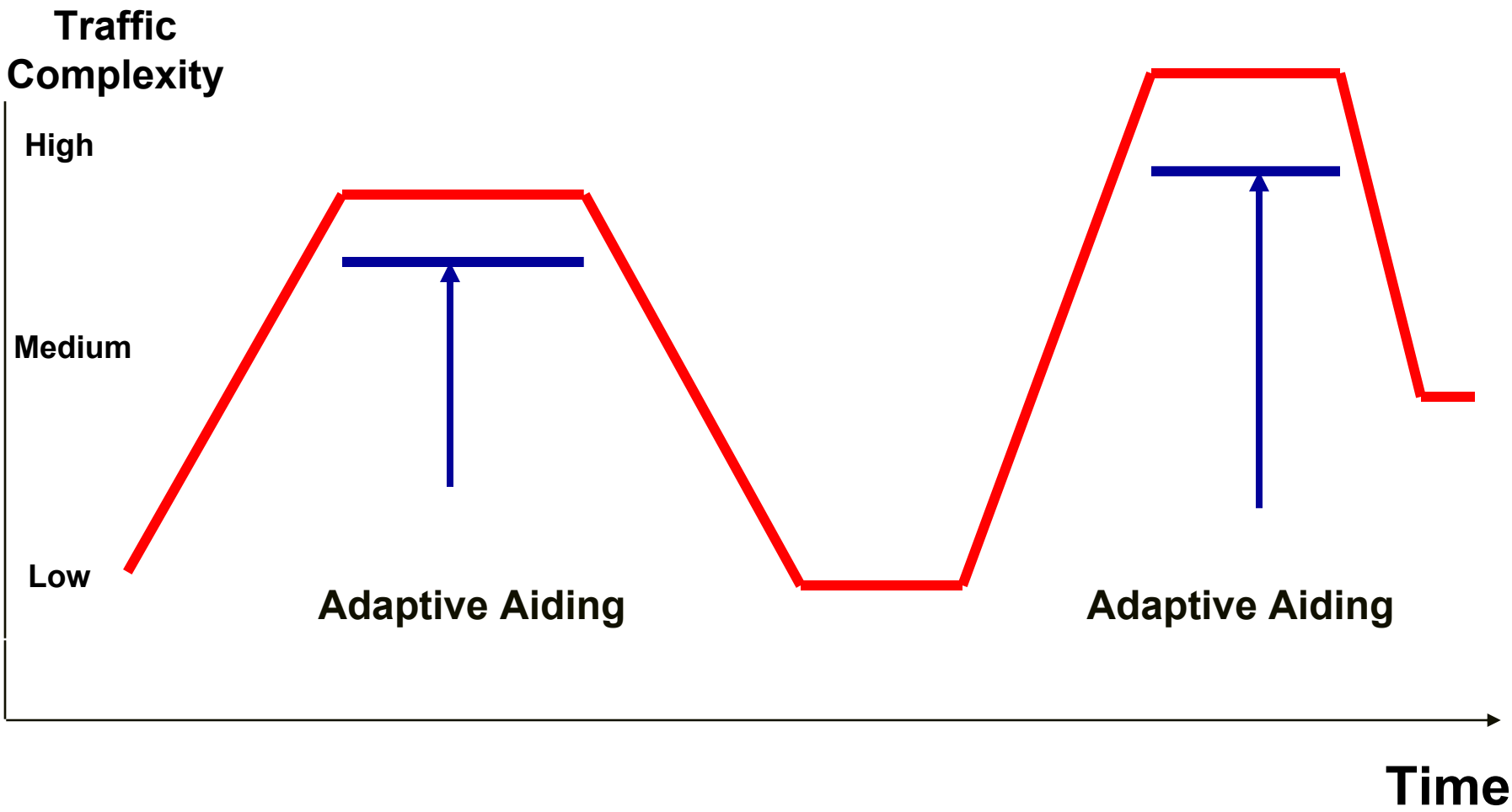


Descent Advisor (DA) advises fuel efficient descent to meet time

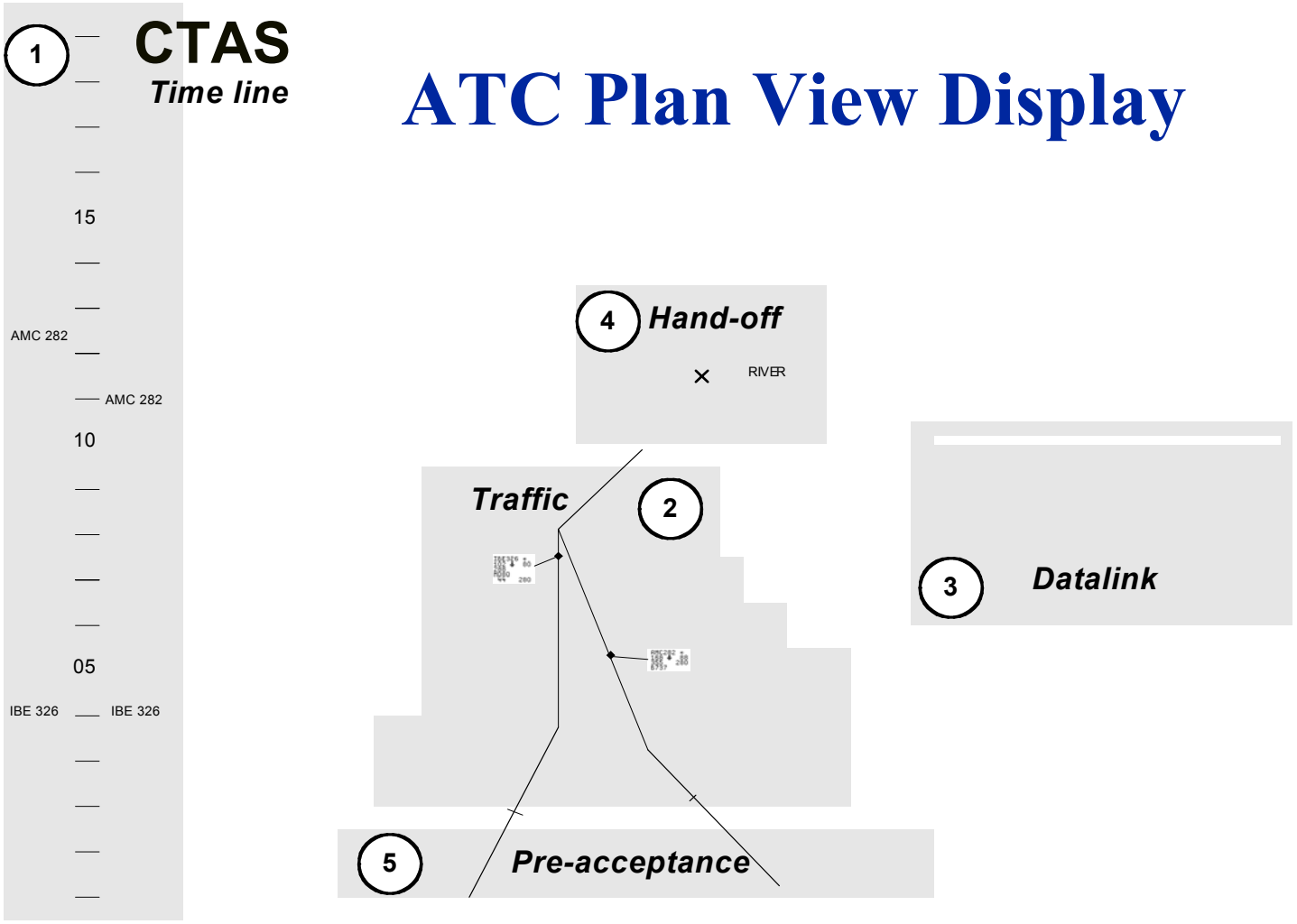




CTAS Adaptive Automation



ATC Plan View Display

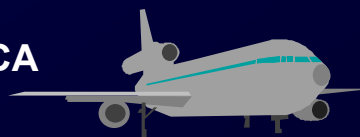


CTAS Aiding

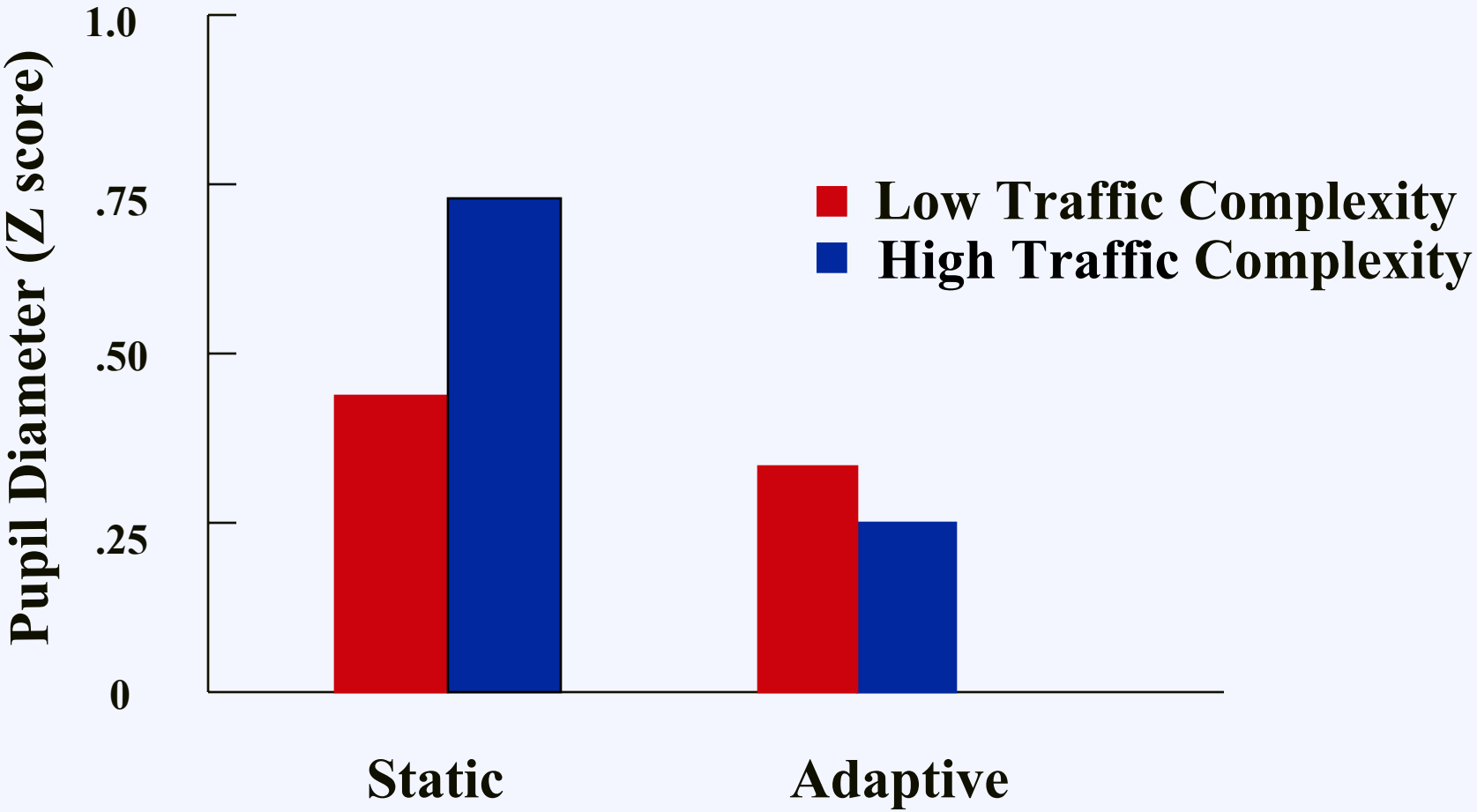
- **Static — CTAS resolution advisories provided throughout scenario irrespective of traffic complexity**

- **Adaptive — CTAS advisories provided during high traffic complexity only**





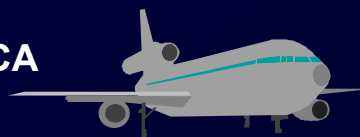
EFFECTS OF ADAPTIVE AUTOMATION AND TRAFFIC LOAD ON AIR TRAFFIC CONTROLLER MENTAL WORKLOAD (Hilburn, Jorna, Byrne, & Parasuraman, In *Human-Automation Interaction,* 1997)



CTAS Adaptive Automation:

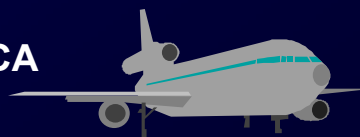
- *Increased Scanning of Primary Traffic*
- *Reduced Mental Workload*
- *Improved Flight Arrival Accuracy*





Benefits of Dynamic Task Allocation

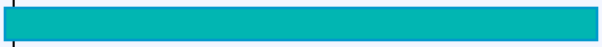




Some Potential Costs of Automation

- Loss of situation awareness
- Mode error/confusion
- Manual skill degradation
- Over-reliance and complacency 
- Unbalanced workload

Multi-Attribute Task Battery (MAT)





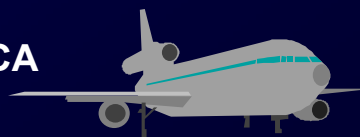


General Methods

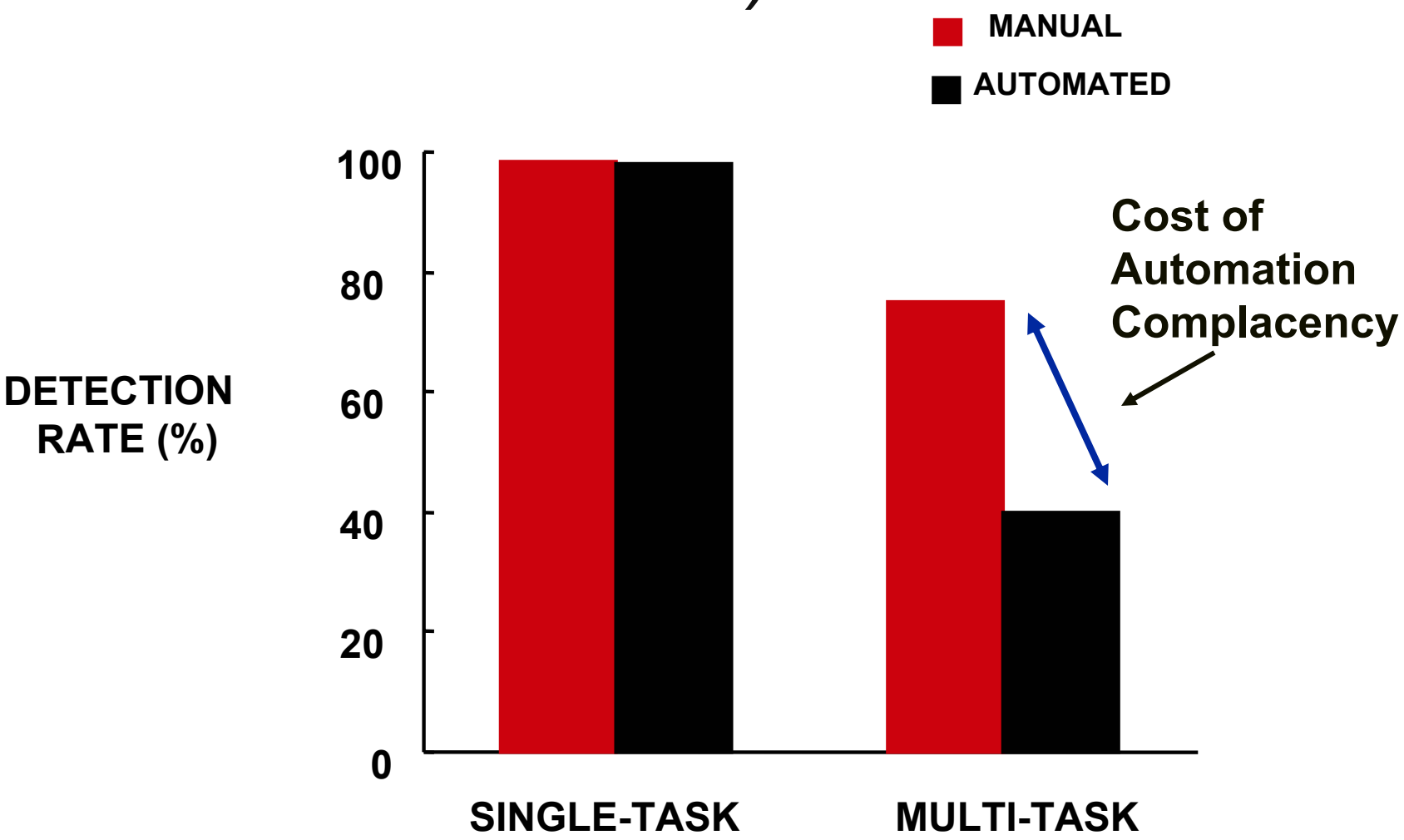
- Subjects: Students and Experienced General Aviation Pilots
- 2 Levels of Difficulty—Single and Multiple-Task
- 2 Levels of Automation (Manual, Automated)

Task: Carry out primary flight and fuel management tasks manually, monitor automated engine-systems task





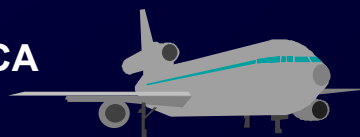
Automation Complacency (Parasuraman, Molloy, & Singh, *International Journal of Aviation Psychology, 1993)*



Cognitive Mechanisms of Automation Complacency

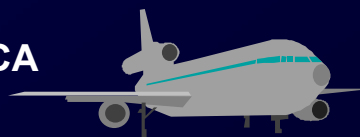
- An Attention Allocation Strategy away from the Automated Task Due to High Trust
- A Decline in Memory for the Dynamics of the Automated Task, Resulting in a Degradation of the Mental Model of the Task

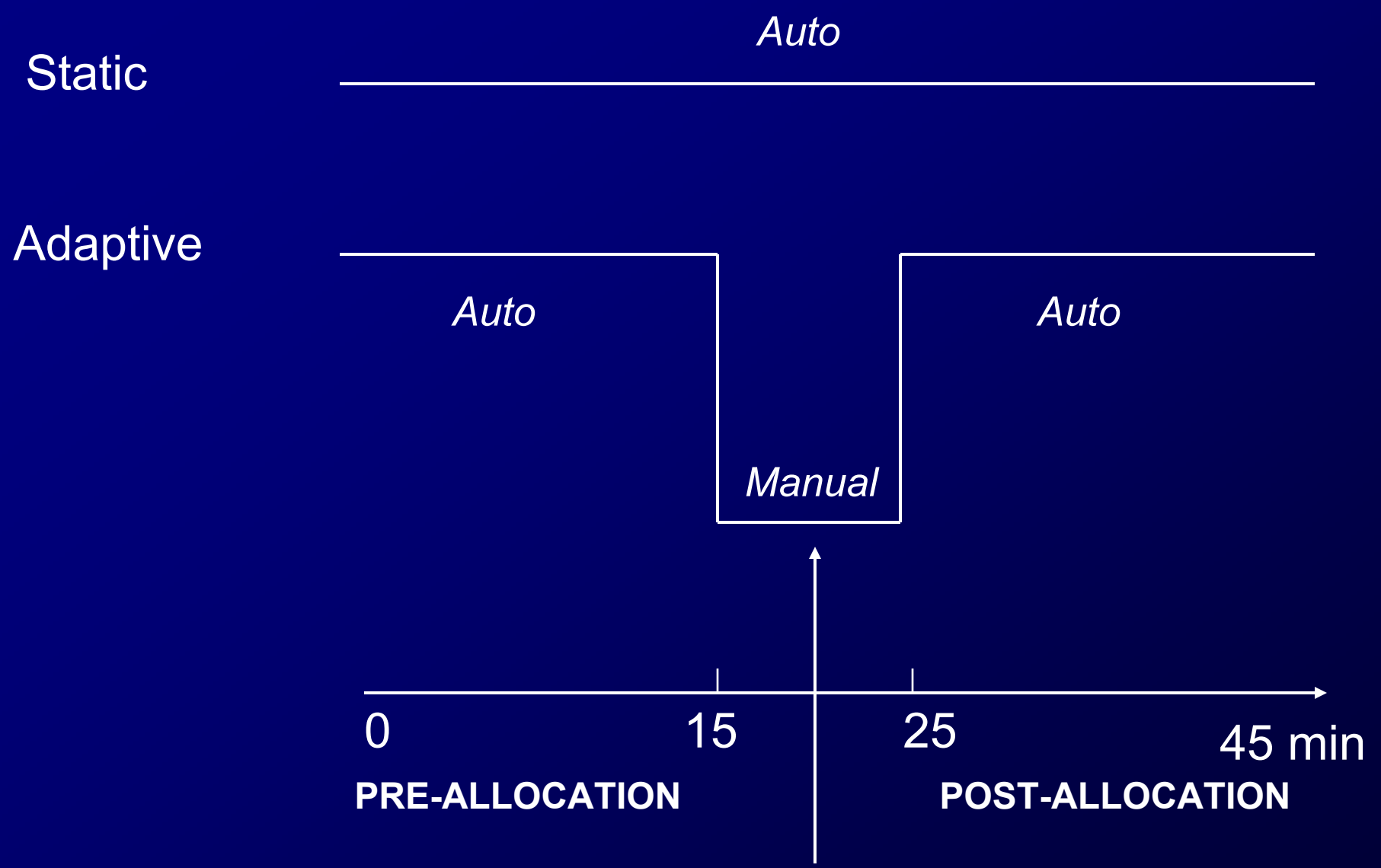




Can Adaptive Task Allocation of an Automated Task Temporarily to Human Control Reduce Automation Complacency?

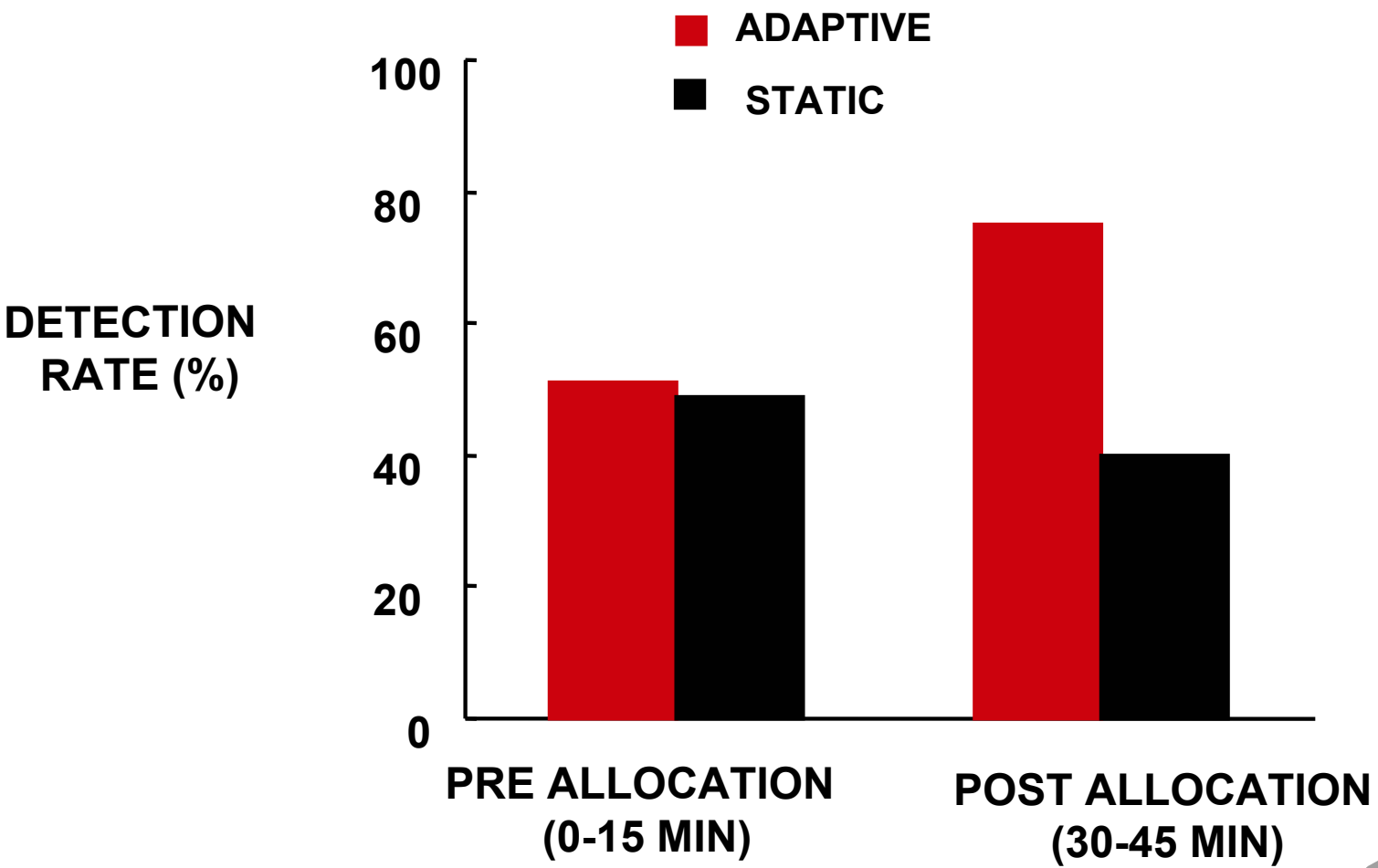




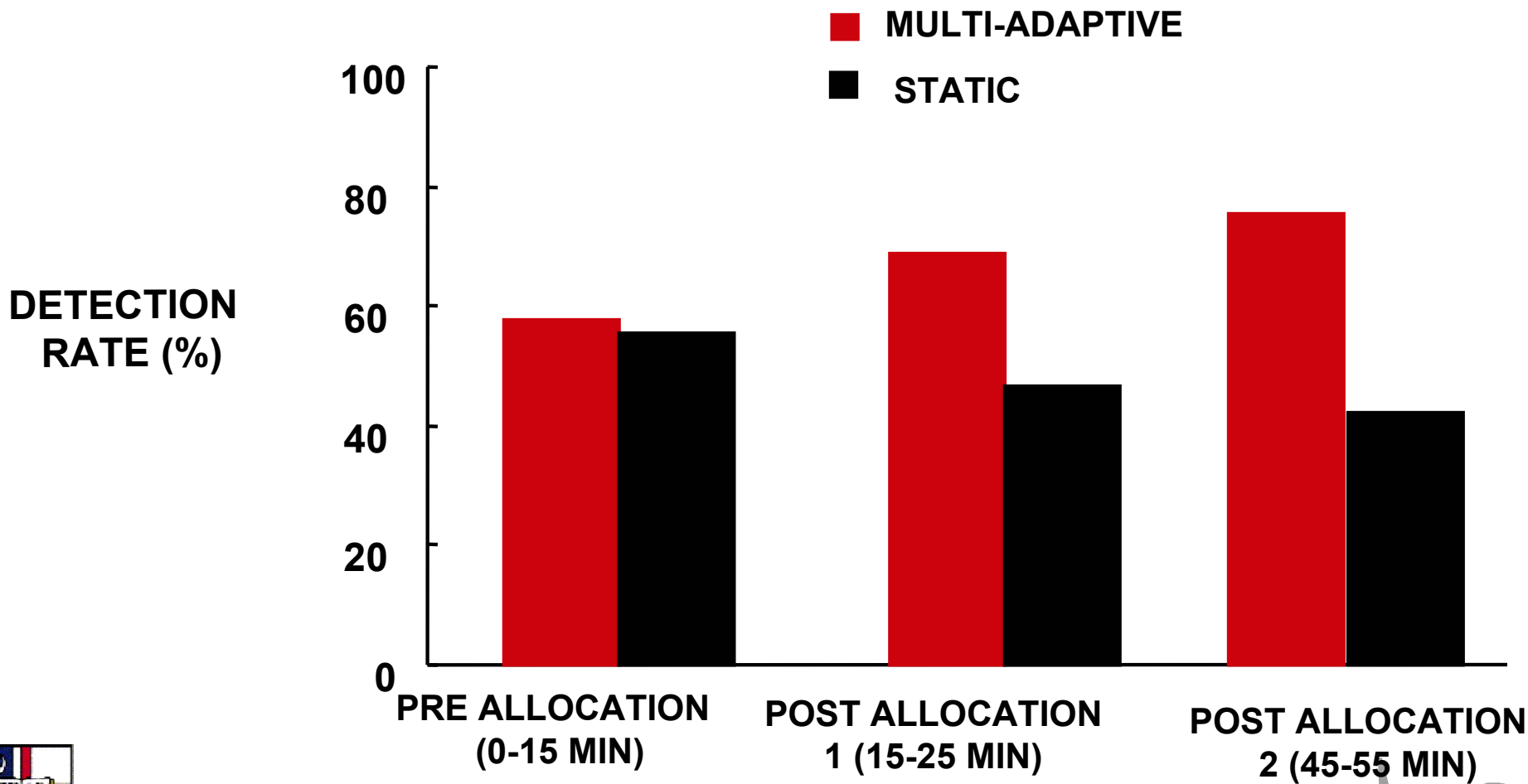


ADAPTIVE TASK ALLOCATION

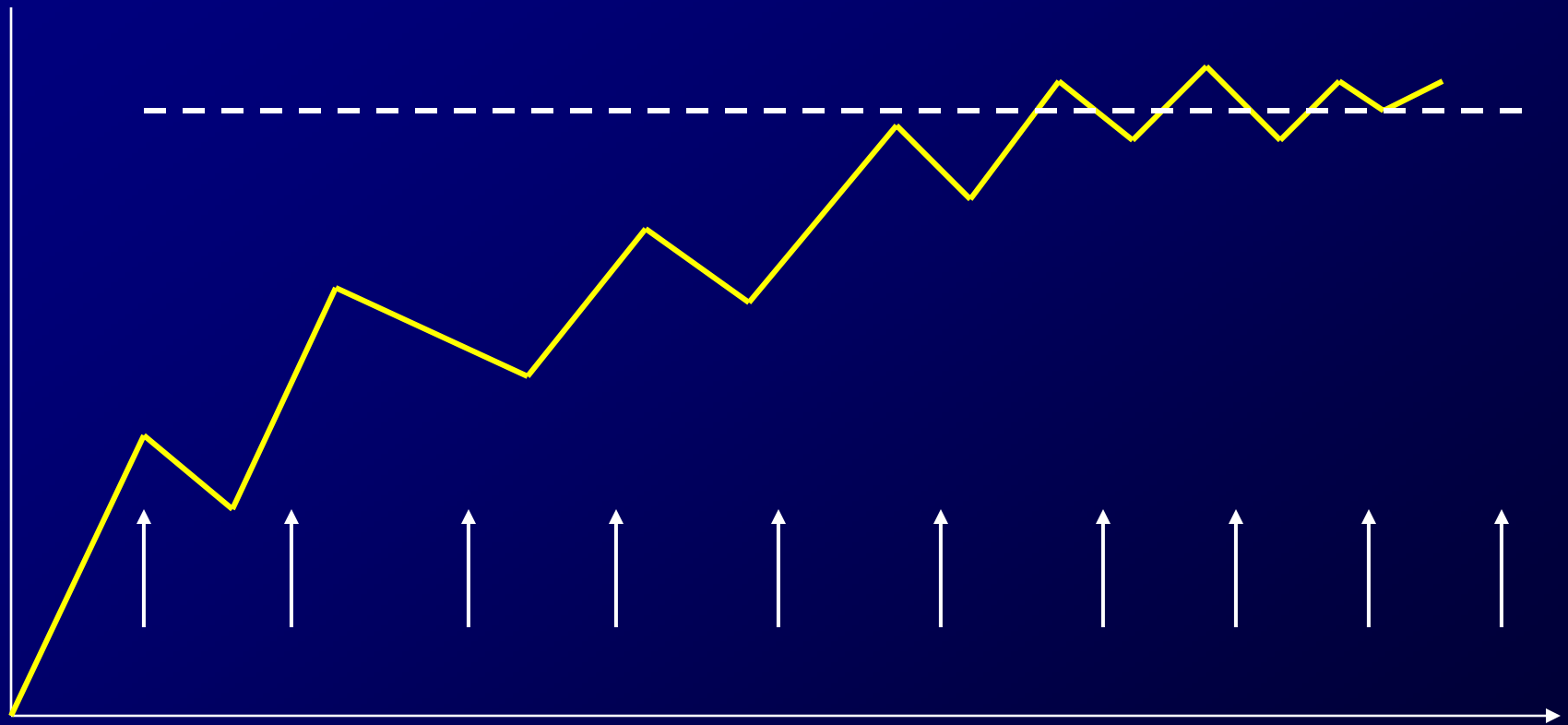
EFFECTS OF ADAPTIVE TASK ALLOCATION
(Parasuraman, Mouloua, & Molloy, *Human Factors, 1996)*



EFFECTS OF MULTIPLE ADAPTIVE TASK ALLOCATION
(Mouloua, Molloy, & Parasuraman, *Proc. Human Factors, 1993*)

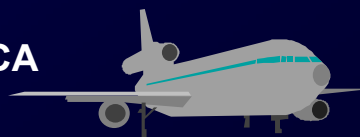


Neural Network Simulation of Repeated Adaptive Task Allocation (Farrell & Lewandowsky, *J. Exp. Psychol: Learn. Mem. Cog.*, 2000)

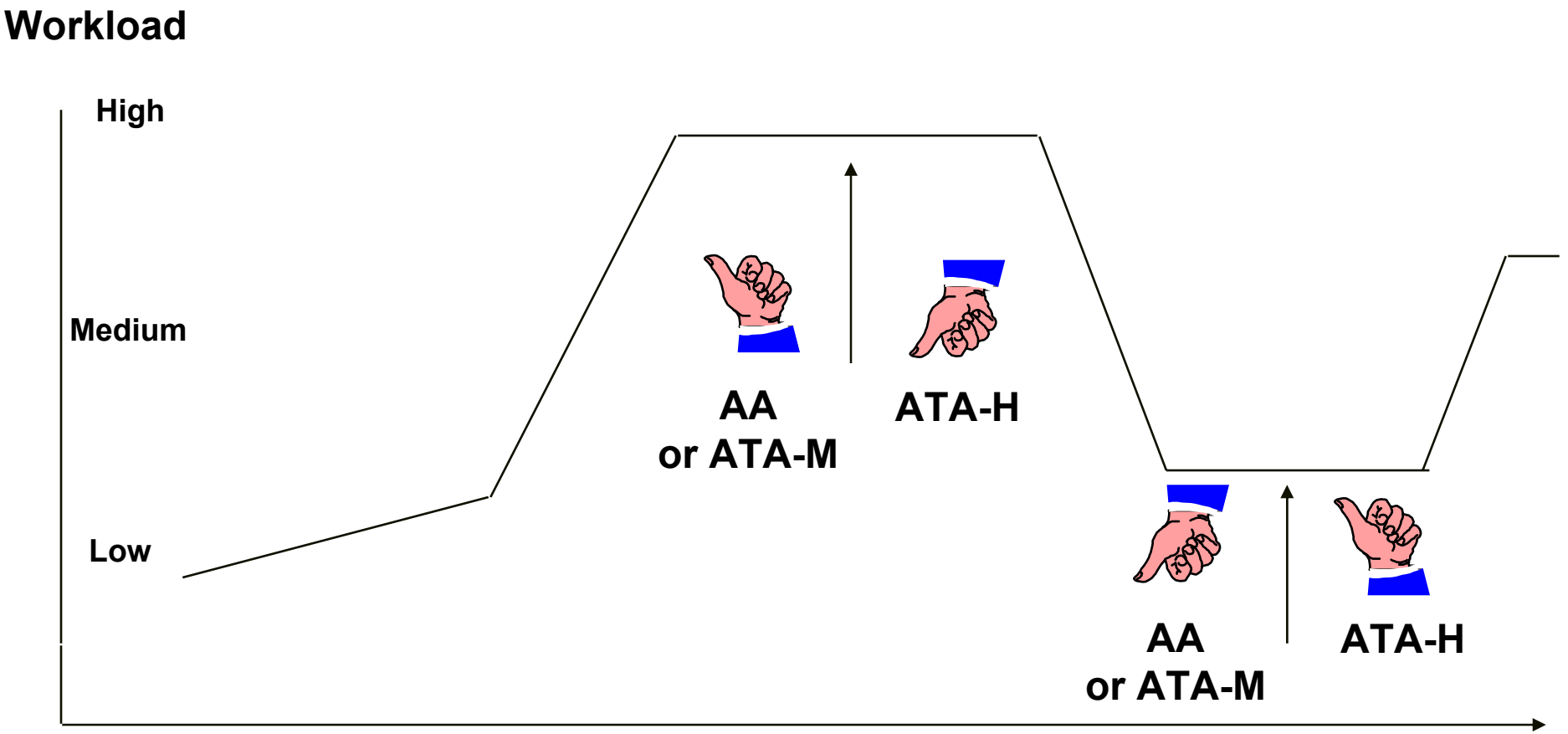


Benefits of Joint Adaptive Aiding and Dynamic Task Allocation





Workload-Matched Adaptive Automation



AA = Adaptive Aiding
ATA-H = Adaptive Task Allocation to Human
ATA-M = Adaptive Task Allocation to Machine

Optimal Time for Adaptation

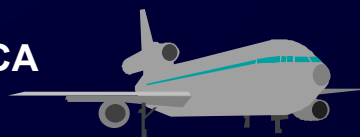
Poor Time for Adaptation

EFFECTS OF WORKLOAD-MATCHED ADAPTIVE AUTOMATION (Parasuraman, Mouloua, & Hilburn, In *Automation Technology & Human Performance,* 1999)

Methods

- Subjects: 24 GA Pilots assigned to 3 groups
- Task: MAT with EICAS Monitoring
- Duration: 60 minutes (3 20-minute blocks)
- Tracking difficulty (bandwidth) varied every 20 minutes: High-Low-High

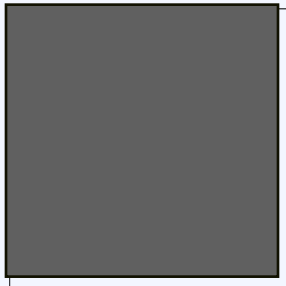




Engine Indicator and Crew Alerting System (EICAS)



Used With the Multi-Attribute Task (MAT)



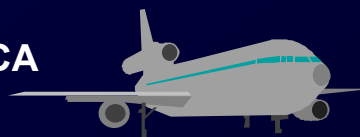
Methods (contd.)

- Adaptive aiding (AA): Altitude hold for tracking task

- Adaptive task allocation to human

  (ATA-H): Allocation of automated EICAS task to manual control for 10 minutes in middle of 20-min block

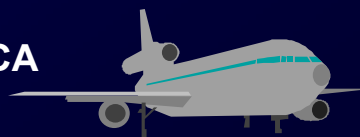




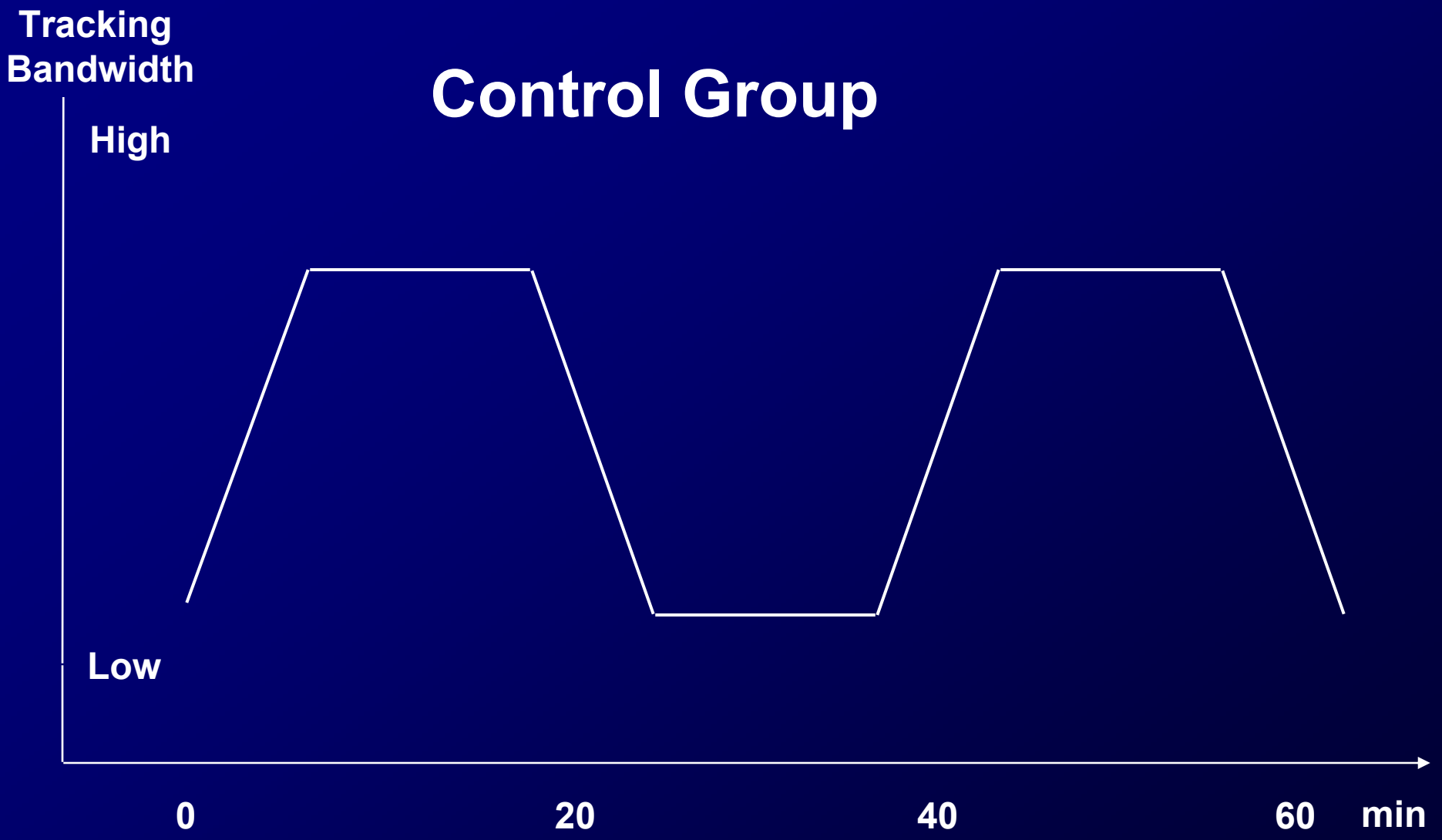
Groups

- Control (Non-adaptive) Group
- Workload-matched Adaptive Group
- Workload-mismatched Adaptive ("Clumsy" Group)

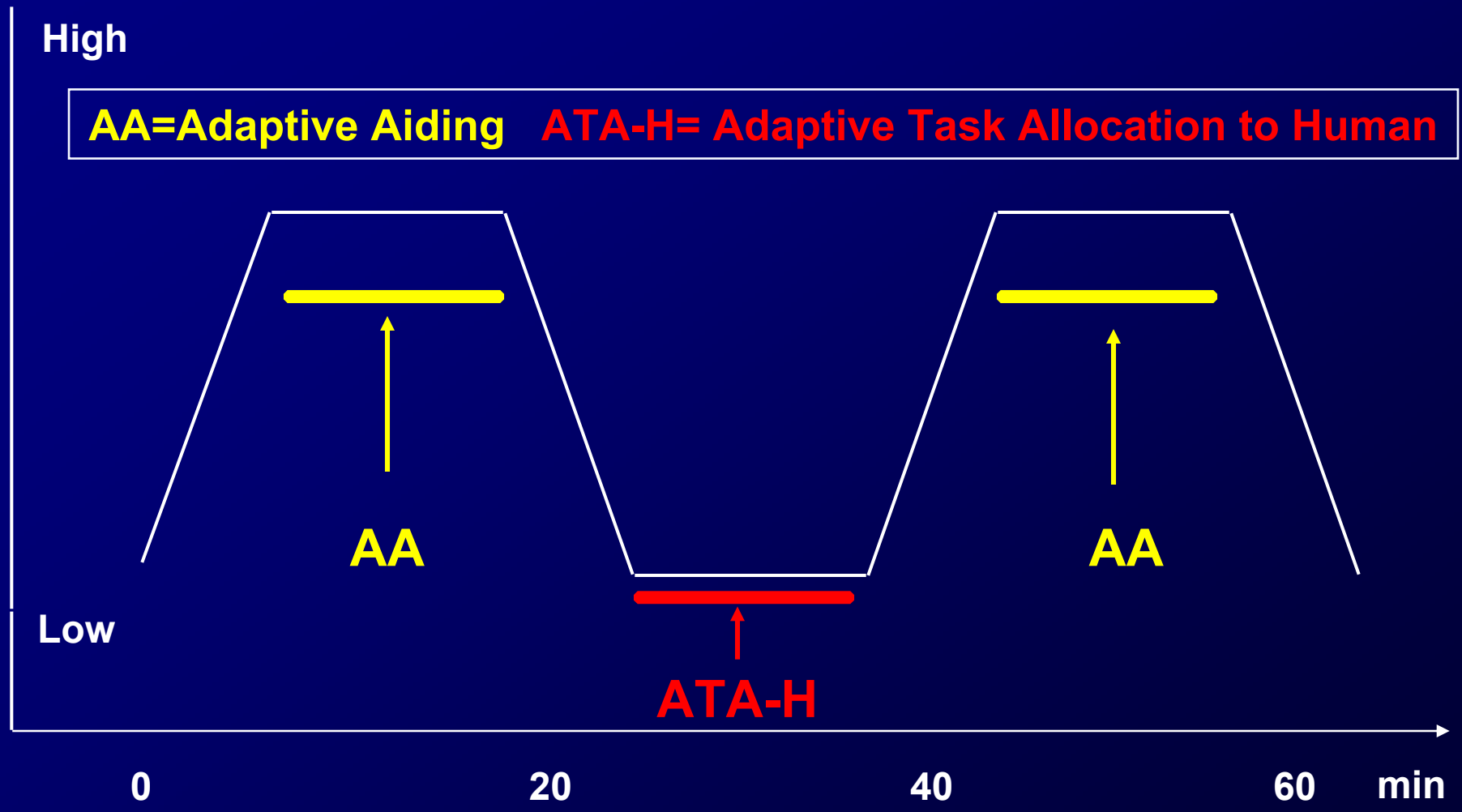




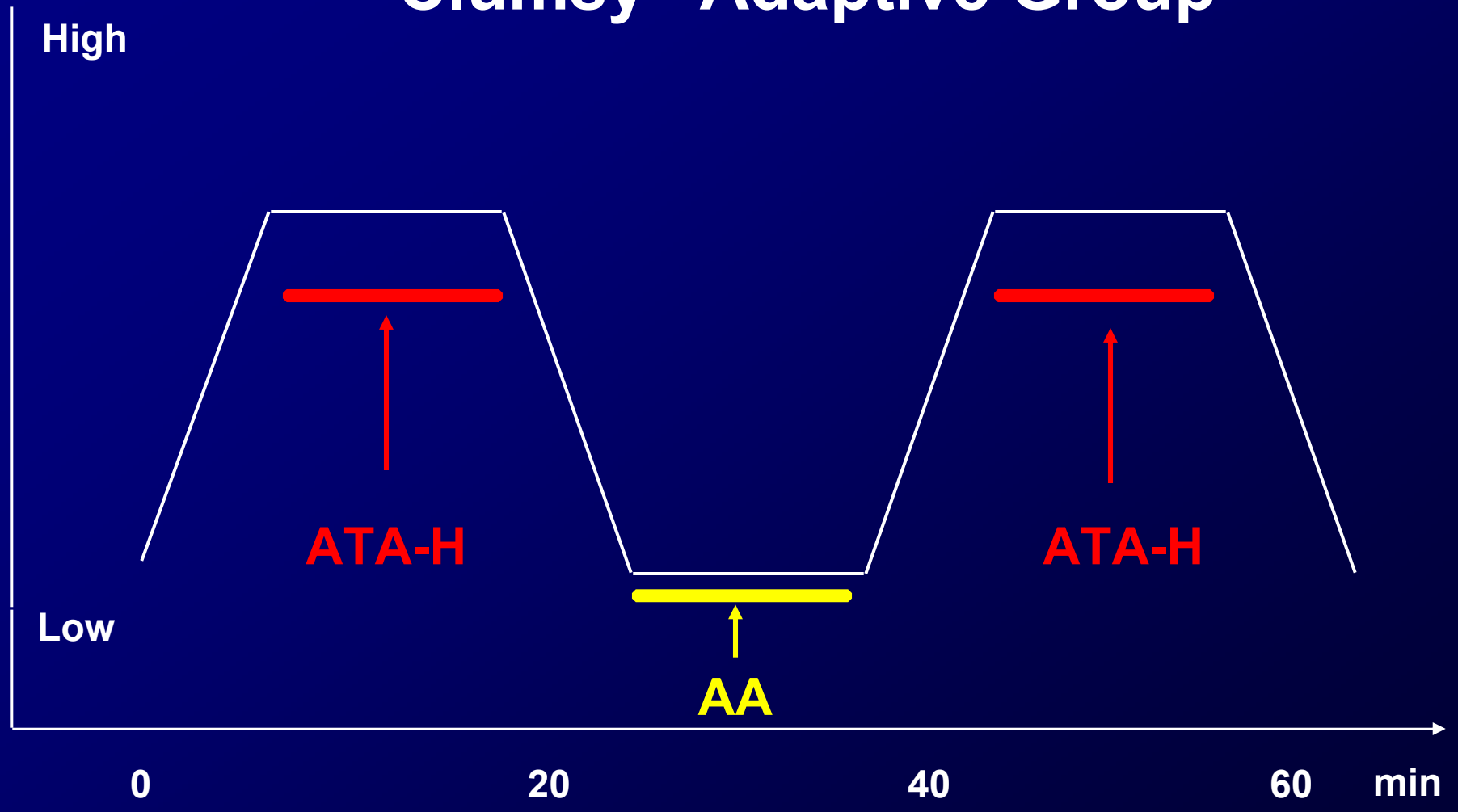
Control Group



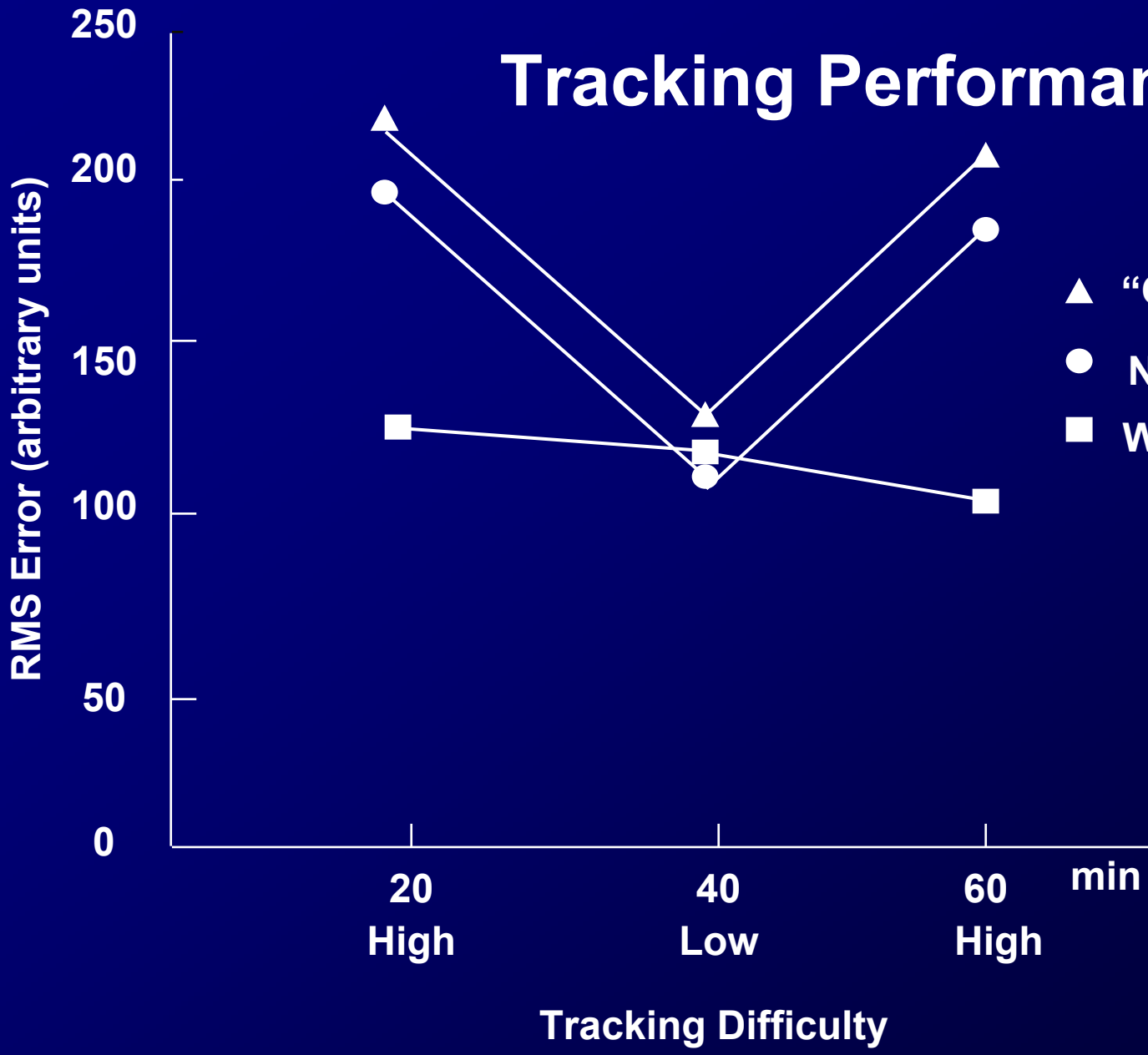
Workload-Matched Adaptive Group



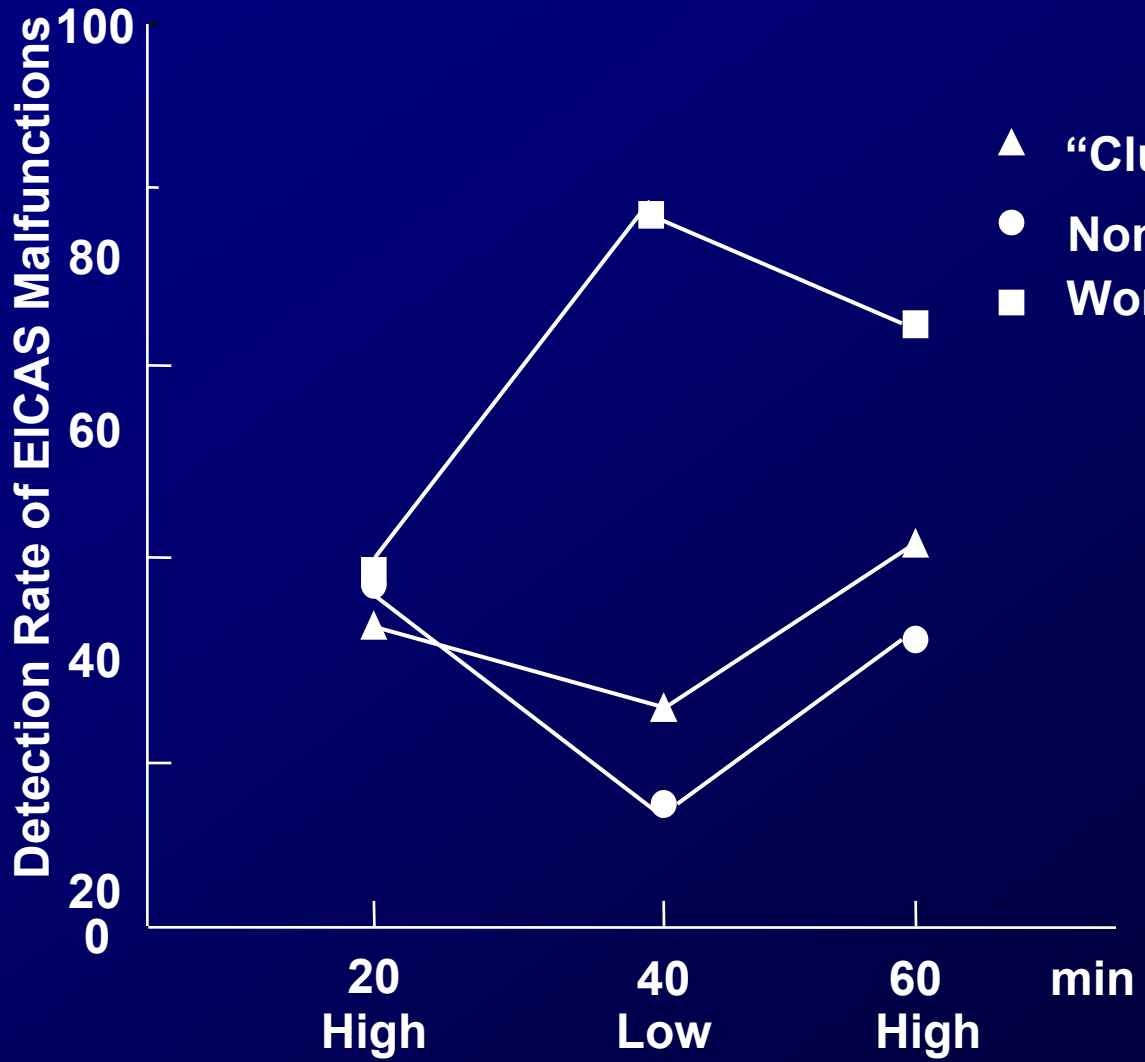
"Clumsy" Adaptive Group



Tracking Performance



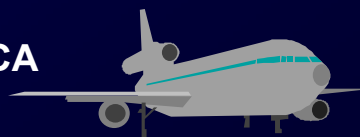
EICAS Malfunction Detection Performance



Interim Conclusion

Adaptive automation matched to human mental workload enhances performance and reduces some of the negative consequences of automation, e.g., complacency, unbalanced workload, and reduced situation awareness

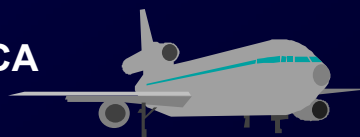




Issues

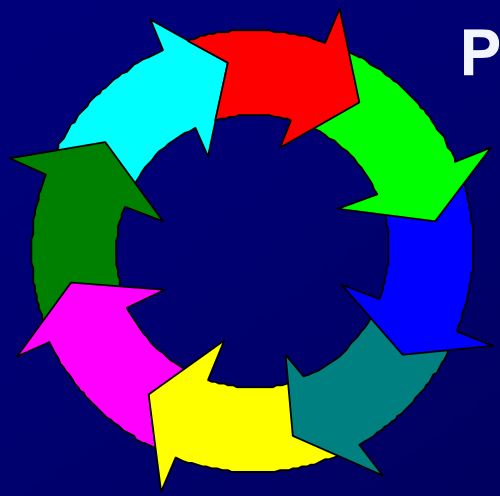
- Is successful adaptation feasible on the basis of *measured* or *predicted* rather than *assumed* workload?
- Performance (behavioral) measures can be scarce in highly-automated complex work domains
- Brain measures can be obtained frequently and in real time
- Can brain measures, alone or in conjunction with performance measures, be used for implementing adaptive automation?





Attentional Resources and Mental Workload: Performance Measures

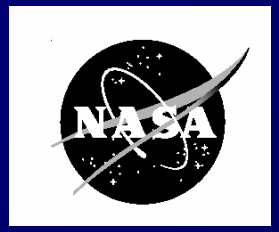
Resources ➲ Performance ⟷ Reaction Time, Accuracy
Performance ➲ Resources



A Problem of Circularity

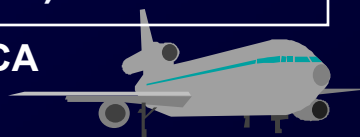
Candidate Physiological Measures for Adaptive Systems

- Heart Rate
- Heart Rate Variability
- Eye Blinks
- Eye Movements
  - Saccades
  - Scan Path Entropy



SOURCE: SCERBO, FREEMAN, MIKULKA, PARASURAMAN, DI NOCERA, & PRINZEL (2001). *The efficacy of psychophysiological measures for implementing adaptive technology.* NASA Technical Report. (NASA/TP-2001-211018).

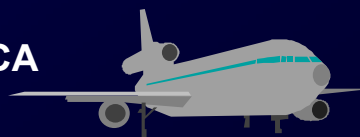




Brain Imaging Techniques

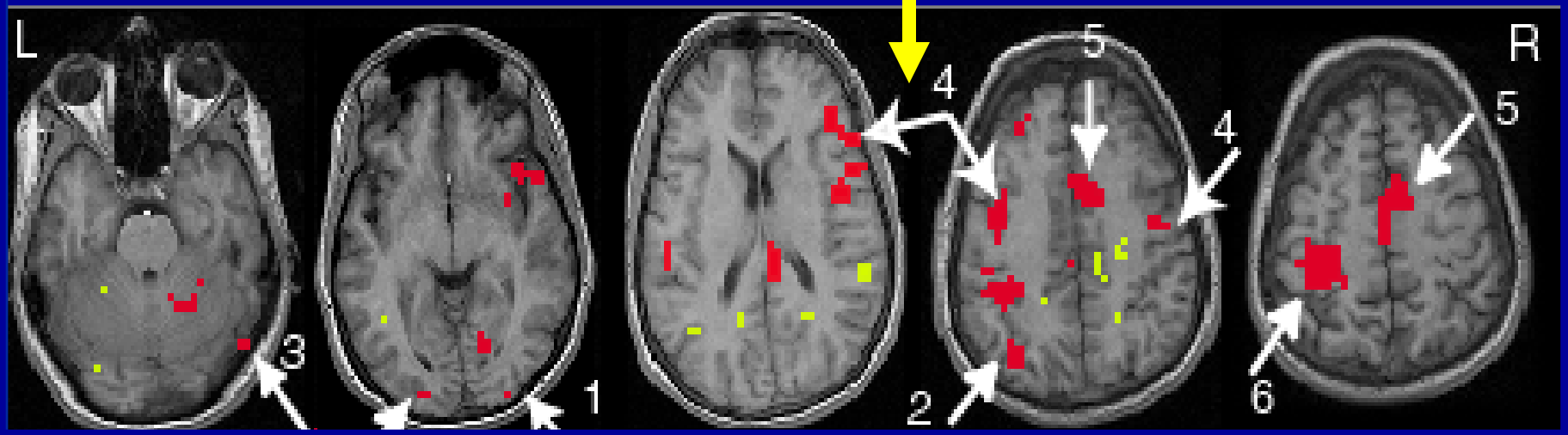
EEG	Electroencephalography
ERPs	Event-Related Potentials
PET	Positron Emission Tomography
fMRI	Functional Magnetic Resonance Imaging
MEG	Magnetoencephalography
TCDS	Transcranial Doppler Sonography
OBI	Optical Brain Imaging



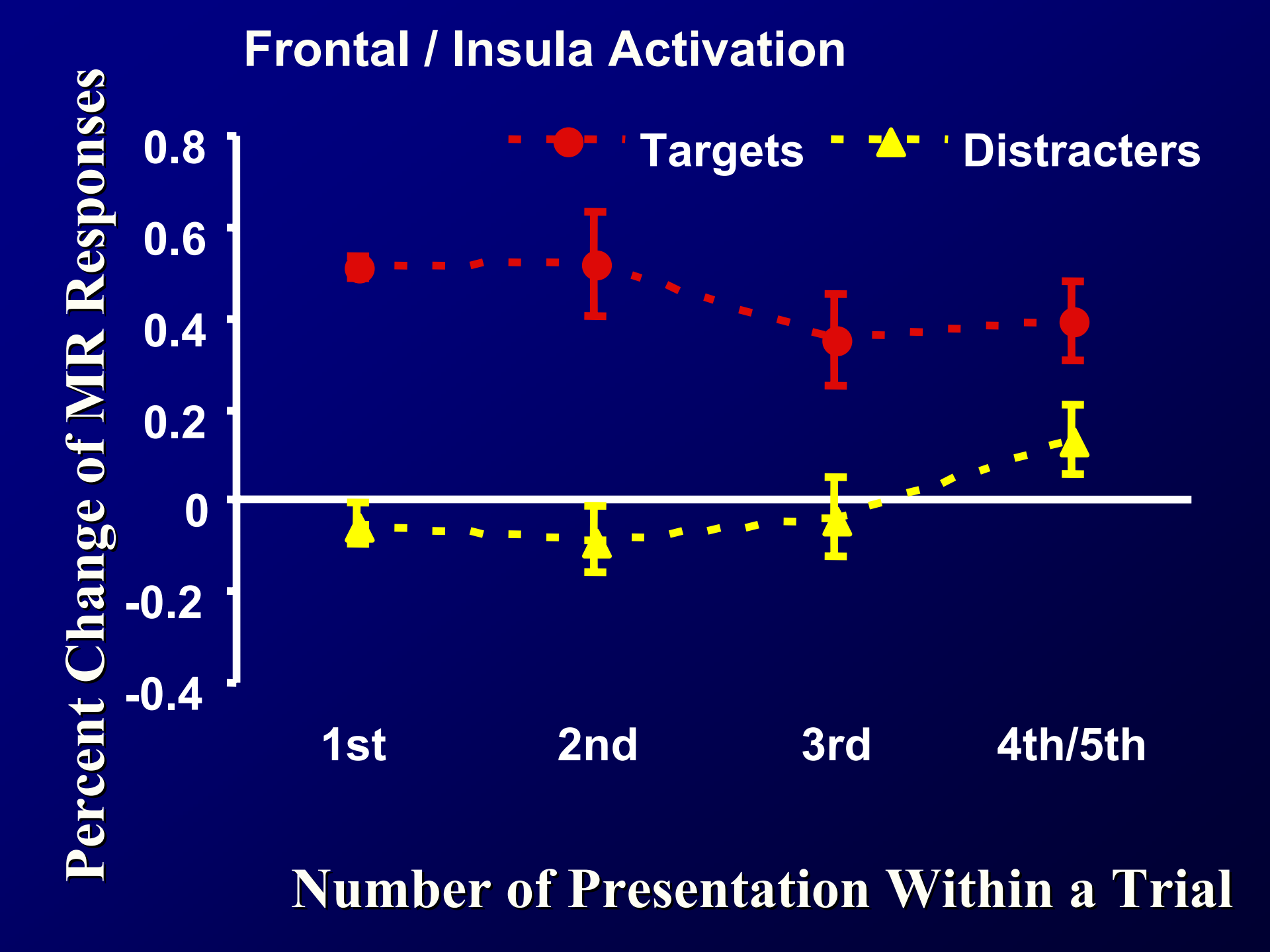


Target-Related Activation During Sustained Working Memory

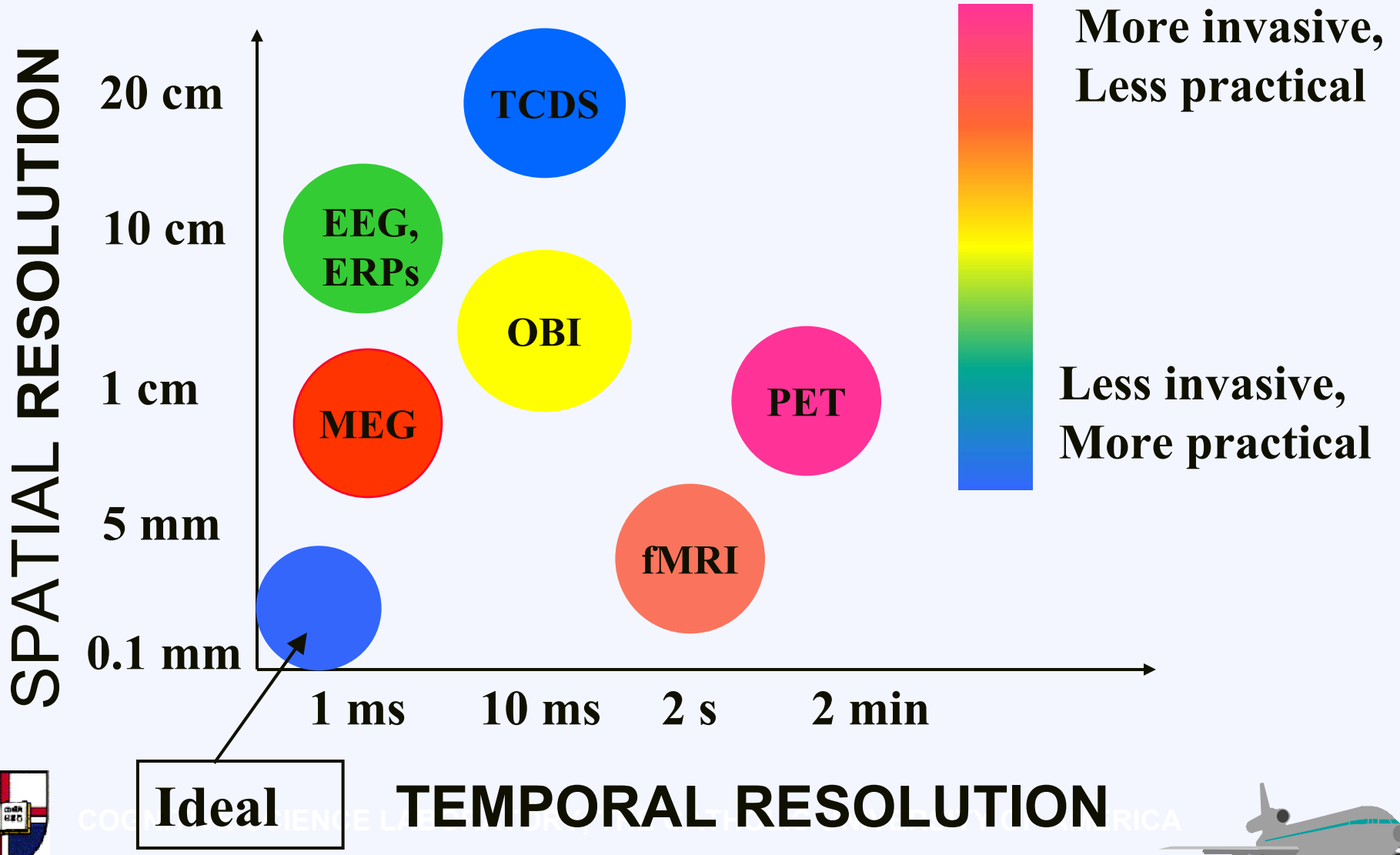
Frontal Cortex/Insula (4, 5)



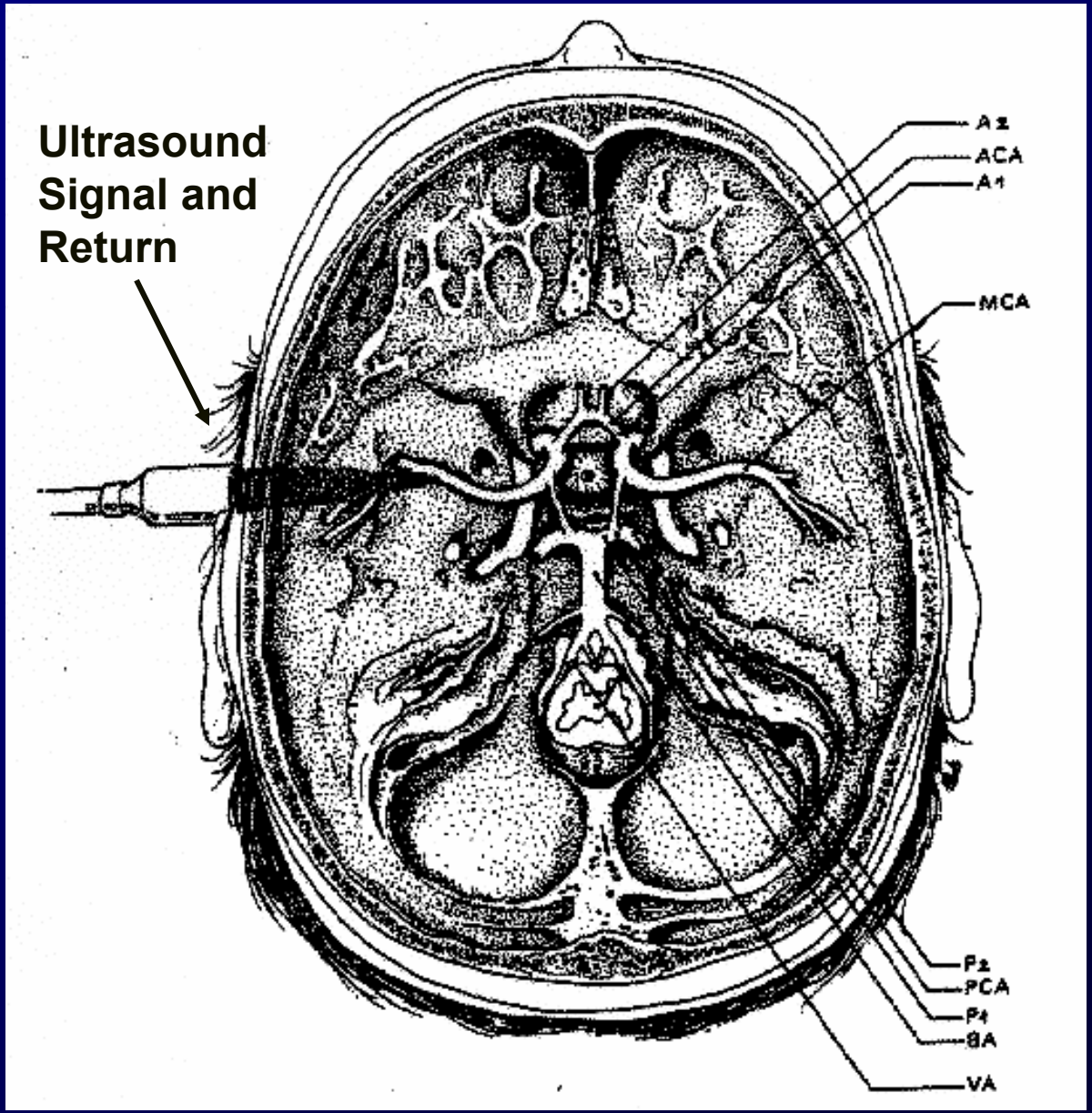
Source: JIANG, Y., HAXBY, J. V., MARTIN, A., UNGERLEIDER, L. G., & PARASURAMAN, R. (2000). Complementary neural mechanisms for tracking items in human working memory. *Science, 287,* 643-646.


Frontal / Insula Activation
Targets
Distracters
Percent Change of MR Responses
0.8
0.6
0.4
0.2
0
-0.2
-0.4
1st
2nd
3rd
4th/5th
Number of Presentation Within a Trial

RESOLUTION SPACE OF BRAIN IMAGING TECHNIQUES FOR ERGONOMIC APPLICATIONS



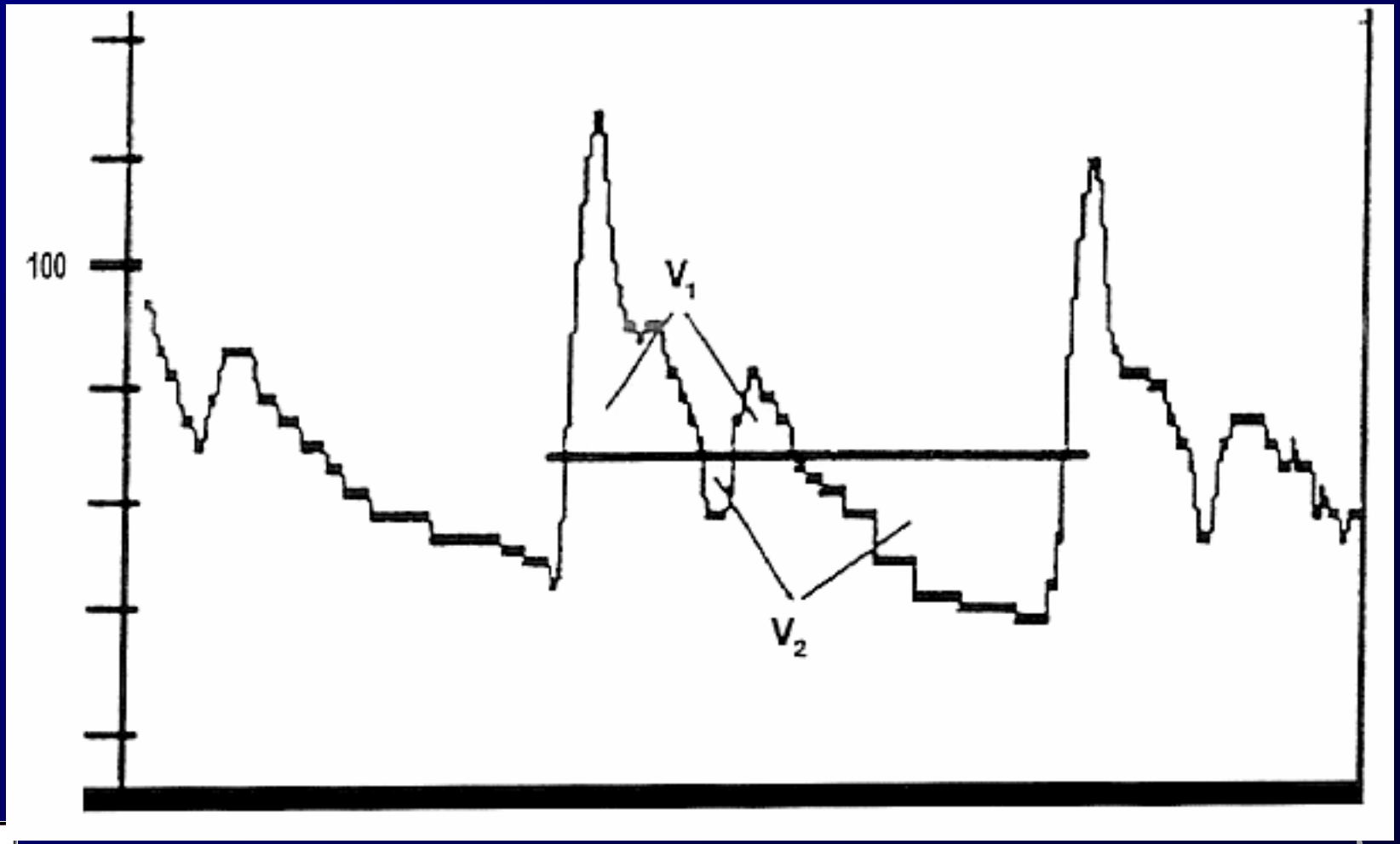
Transcranial Doppler Sonography (TCDS)



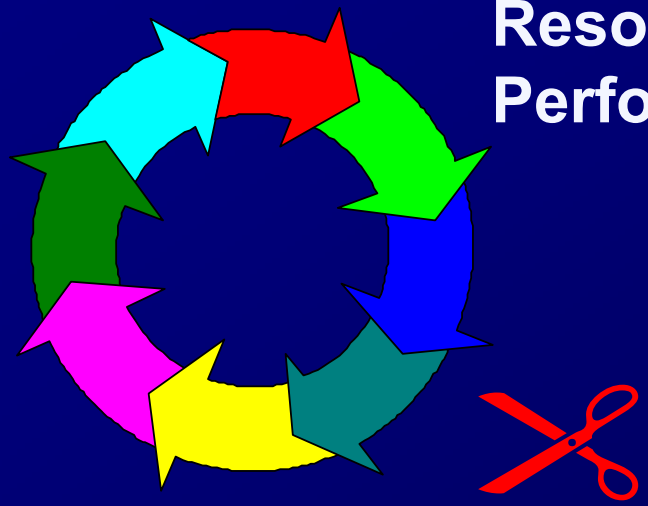
Transcranial Doppler Sonography System (TCDS)



TCDS Measured Cerebral Blood Flow Velocity



Attentional Resources: Breaking Out of the Circularity



Resources ➲ Performance
Performance ➲ Resources

⟷ Reaction Time, Accuracy

Cortical Activation Measures ⟷



Neuroergonomic Measurement of Attentional Resources and Mental Workload

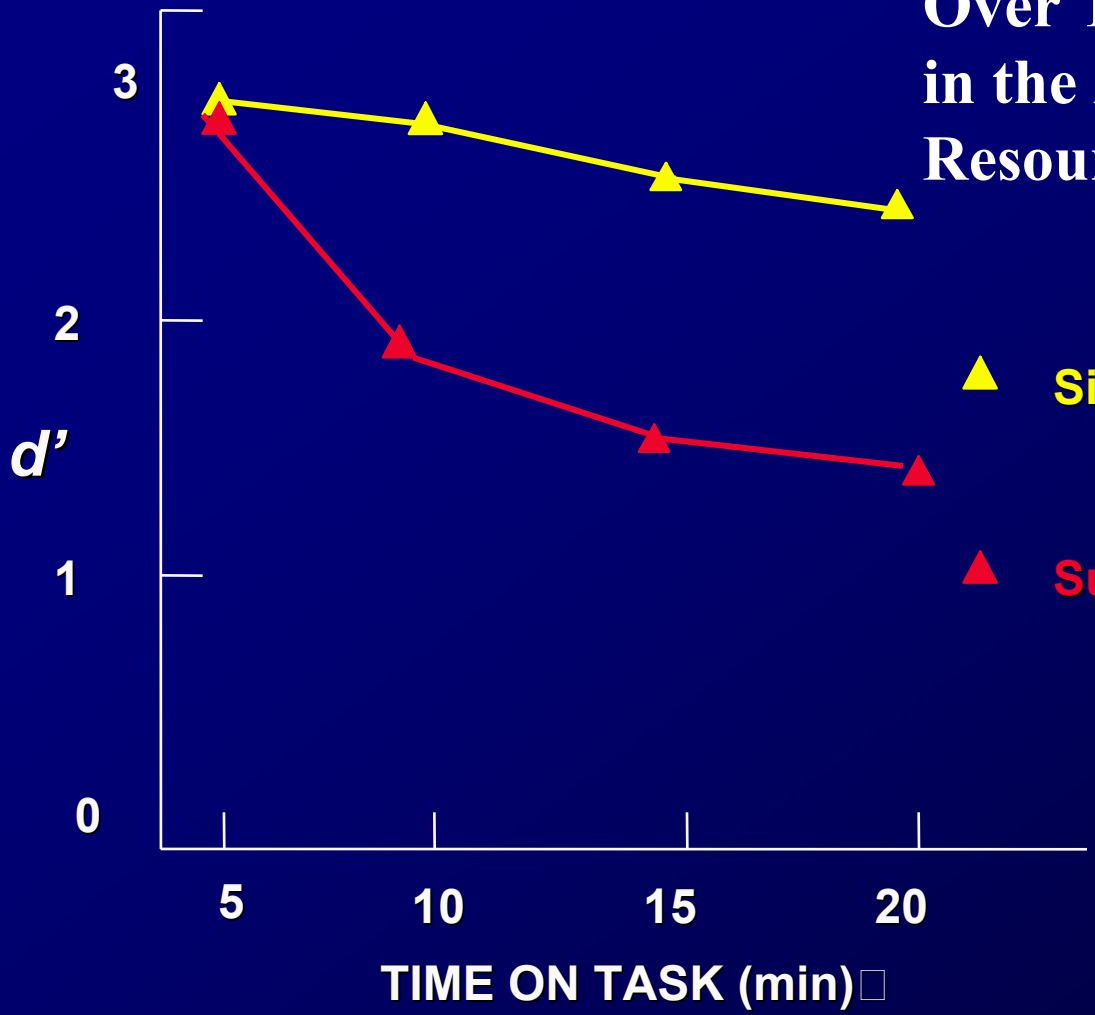
- Cortical Activation and Blood Flow (fMRI and OBI)
- Cerebral Blood Flow Velocity (TCDS)



Sir Charles Sherrington
(1857-1952)

Roy, C. S., and Sherrington, C. S. (1890). On the regulation of the blood supply of the brain. *Journal of Physiology (London), 11,* 85-108.

Does the Vigilance Decrement Over Time Reflect a Decrease in the Allocation of Attentional Resources to Target Detection?

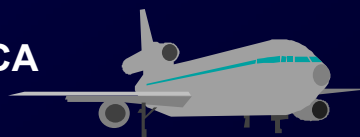


Source: PARASURAMAN, R. (1979). Memory load and event rate control sensitivity decrement in sustained attention. *Science, 205,* 224-227.

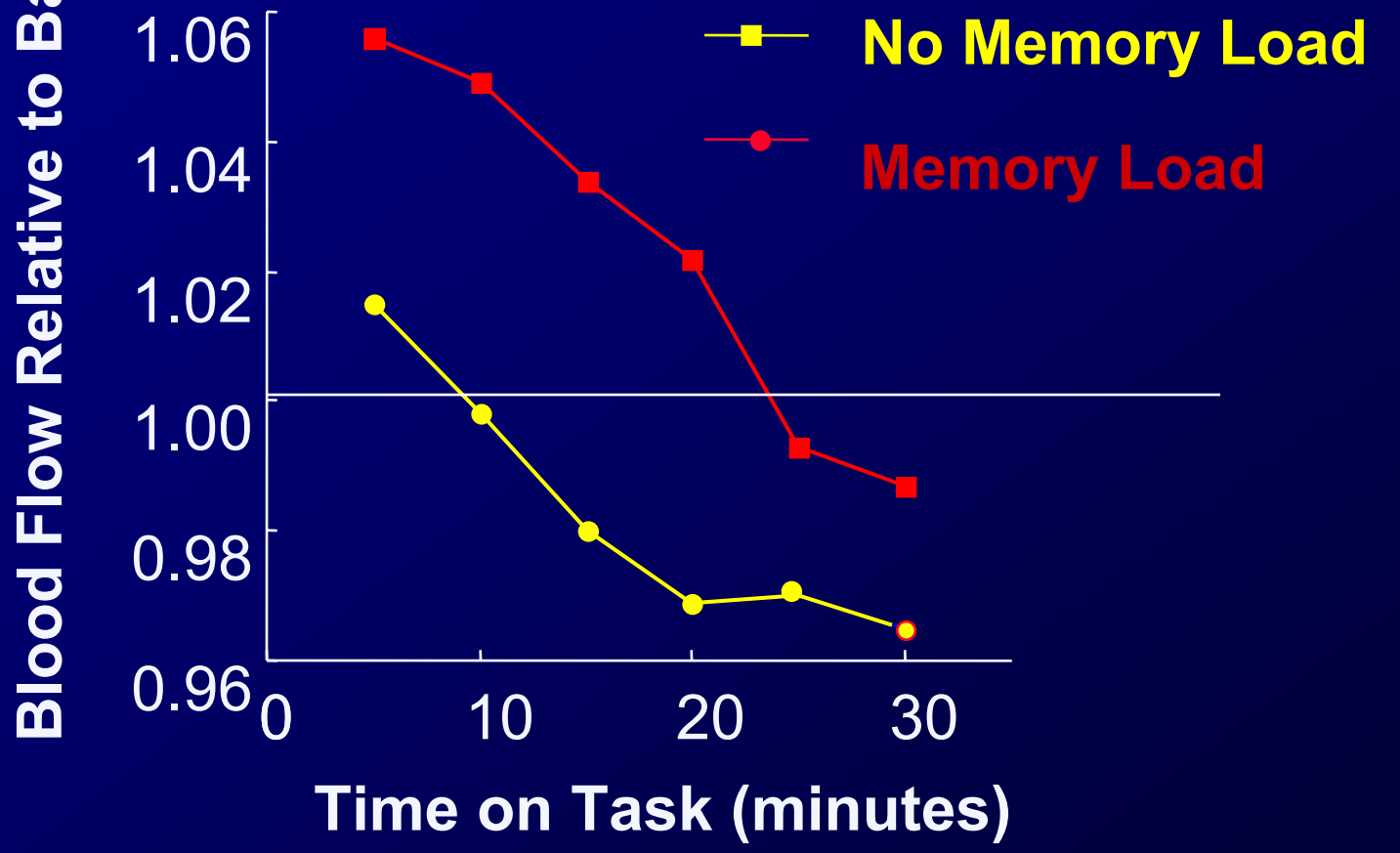
Methods

- **Subjects: 40 Healthy Young Adults**
- **Tasks: Simultaneous and Successive Vigilance (30 Min each)**
- **TCDS Parameters:** Nicolet EME, 2MHz; MCA imaged at 50-55 mm; Time averaged blood flow velocity/3 seconds

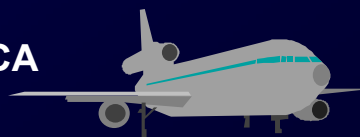




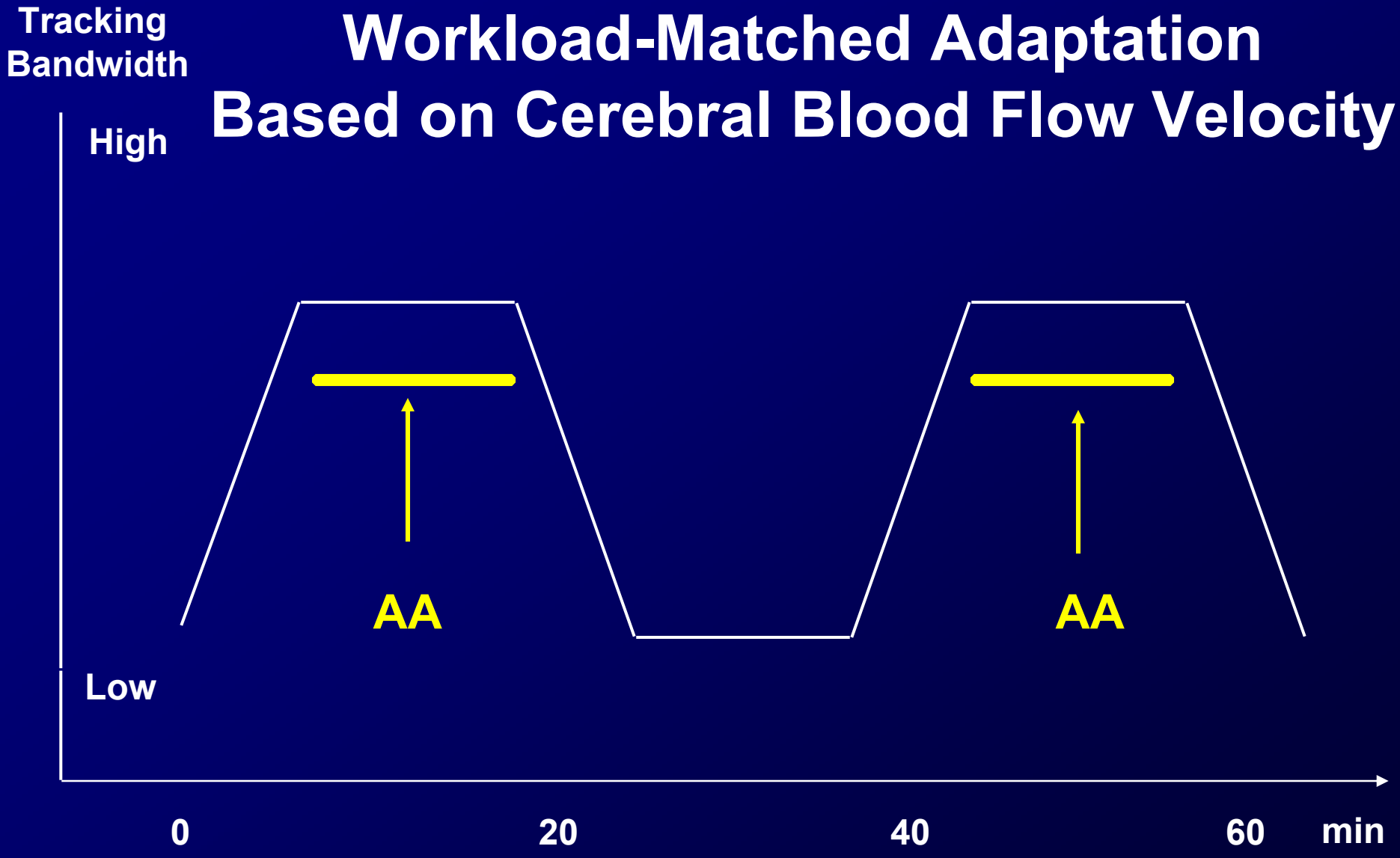
Mayleben et al. (1998)







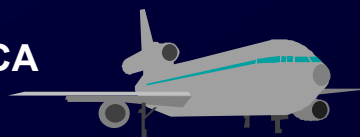
Workload-Matched Adaptation Based on Cerebral Blood Flow Velocity



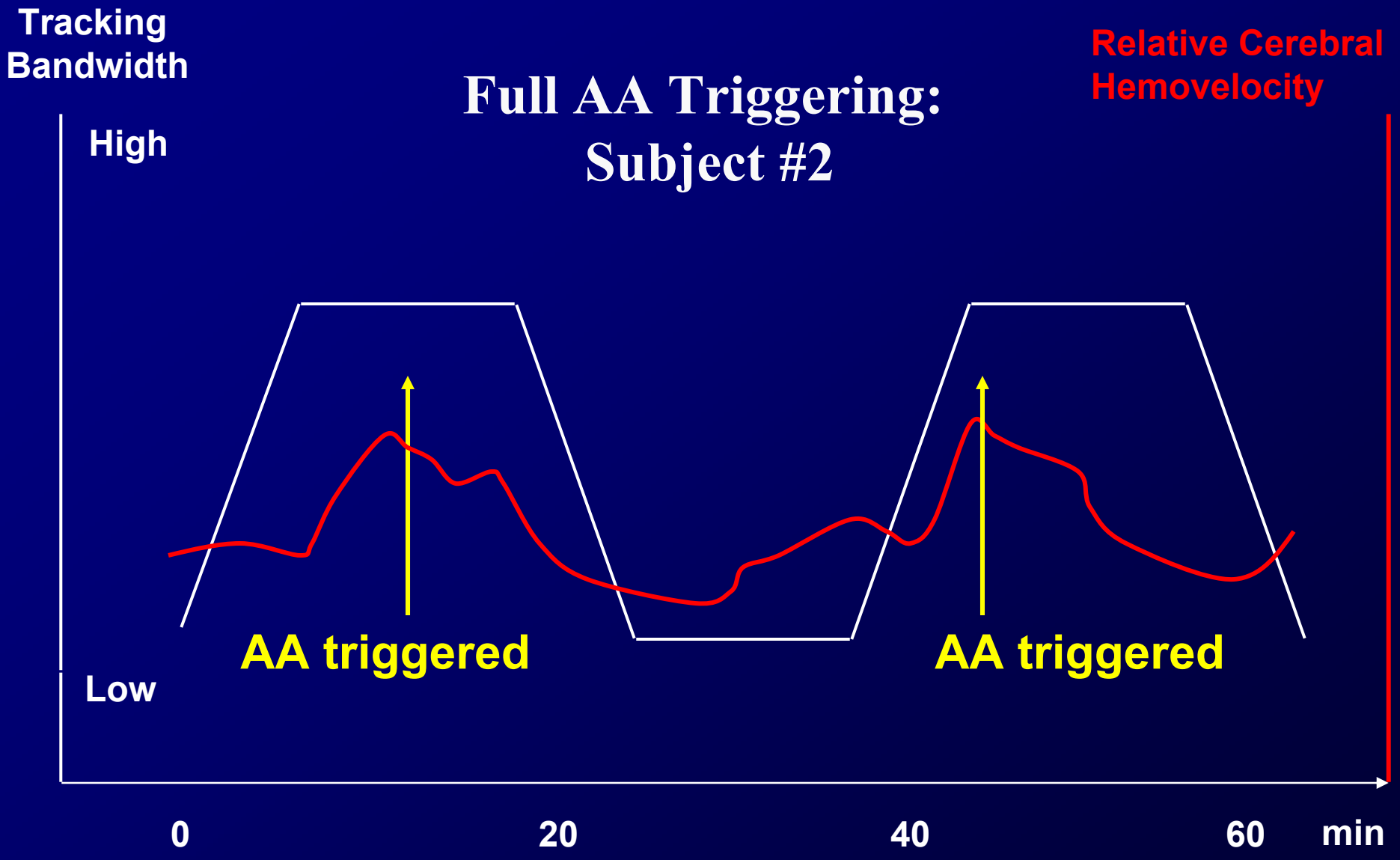
Methods

- **TCDS Parameters:** Nicolet EME, 2MHz; Bilateral MCA imaged at 50-55 mm; Time averaged blood flow velocity/3 seconds
- **60-second moving window of TCDS-measured cerebral hemovelocity**
- **Adaptive aiding triggered when hemovelocity algorithm exceeds 1 SE of operator's baseline**

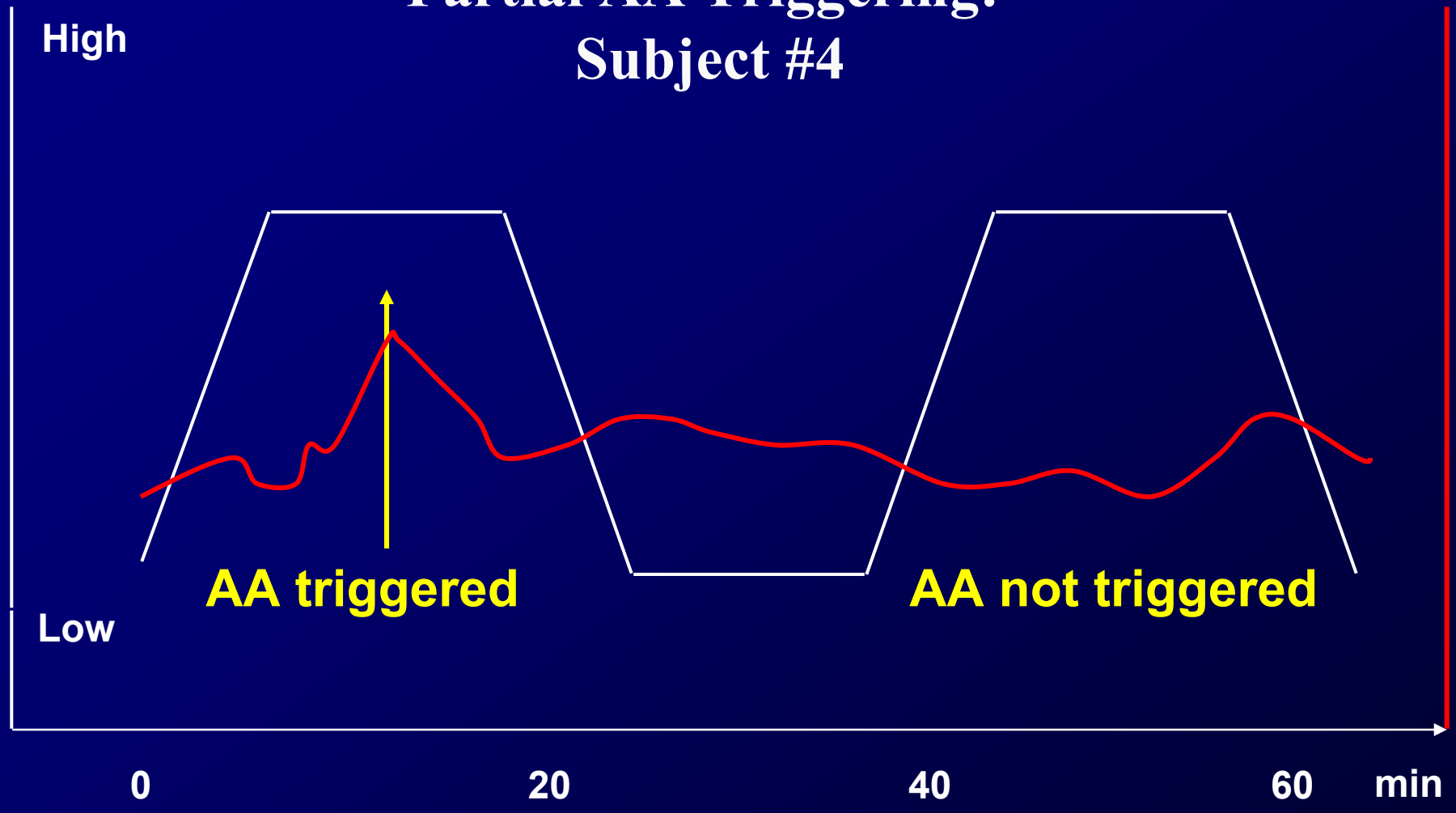




Full AA Triggering: Subject #2

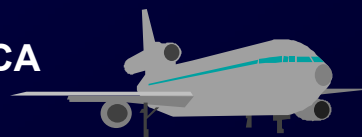


Partial AA Triggering: Subject #4




Tracking Bandwidth
Relative Cerebral Hemovelocity
High
No AA Triggering:
Subject #12
AA not triggered
AA not triggered
Low
0
20
40
60
min

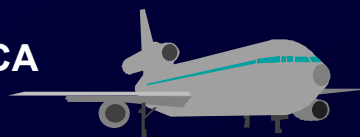




Conclusions

- Real-time adaptive automation matched to human mental workload is feasible with on-line measurement of cerebral blood flow
- Physiologically-driven adaptive automation enhances operator performance and regulates mental workload





The Catholic University Of America
Cognitive Science Laboratory

* Devoted to Studies in Cognitive Neuroscience and Human Factors *

250 O'Boyle Hall
The Catholic University of America
Washington, DC 20064
Office: (202) 319-5825 Fax: (202) 319-4456
e-mail: cua-csl@cua.edu
http://psychology.cua.edu/csl/



Acknowledgments

- *NASA Ames Research Center*
- *NASA Langley Research Center*
- *NASA Goddard Space Research Center*
- *Army Research Institute*

Quarterly Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield

Army MURI, Subcontract to U. Central Florida

Raja Parasuraman
Catholic University of America

PERFORMANCE PERIOD: 3/31/02- 6/30/02

During this period, data analyses were completed for two studies to be presented at the Annual Conference of the Human Factors and Ergonomics Society, Baltimore, MD. In addition, a paper on the use of adaptive automation to counter stress-related performance decrements was presented at the NATO Advanced Research Workshop on Operator Functional States, Il Ciocco, Tuscany, Italy.

Paper Presentations

1. *Ericka Rovira, Kathleen McGarry, & Raja Parasuraman*, Effects of unreliable automation on decision making in command and control. Paper to be presented at HFES, Baltimore, MD, September 30-October 4, 2002.

The effectiveness of automated decision aids used by human operators in command and control systems may depend not only on automation reliability, but also on the type (stage) and level the automated support provides. Automation can be applied to information acquisition, information integration and analysis, decision choice selection, or action implementation. The present study examined the effects of variations in the stage of automation support on performance in a "Sensor to Shooter" targeting simulation of command and control. Independent variables included the type and level of automation support (complete listing, priority listing, top choices, and recommendation of decision choice) and the reliability of the automation (60% and 80%). Dependent variables included accuracy and reaction time of target engagement decisions. Compared to manual performance, reliable automation did not affect the accuracy of target engagement decisions but did significantly reduce decision times. When the automation was unreliable, under the higher reliability condition (80%) there was a greater cost in accuracy performance for higher levels of automation aiding (priority listing, top choice, and recommendation) than at a lower level (complete listing). The results support the view that automation unreliability has a greater performance cost for decision automation than for information automation. This performance cost generalizes across a number of different forms of decision-aiding.

2. Ericka Rovira, Marla Zinni, & Raja Parasuraman, Effects of information and decision automation on multi-task performance. Paper to be presented at HFES, Baltimore, MD, September 30-October 4, 2002.

Automation purported to assist human operators may itself be an additional source of complexity and uncertainty. Because high reliability cannot always be assured, imperfect automation can add to uncertainty and thereby degrade performance. The present study examined the relative benefits and costs of information and decision automation and investigated the effects of uncertainty resulting from automation unreliability during multiple task performance. Subjects were either provided with status information ("information" automation) or a recommendation for action ("decision" automation) for the system monitoring sub-task of the Multi-Attribute Task Battery (MAT). Two levels of automation reliability were compared. The detrimental effect of unreliable automation—a decrease in the detection rate of malfunctions—was greater for automation of higher reliability, a result consistent with previous findings of automation-related complacency. This effect of automation unreliability was also greater for decision than for information automation.

3. *Raja Parasuraman, R.* (2002, April). Adaptive automation matched to human mental workload. Invited Keynote Paper, NATO Advanced Research Workshop on Operator Functional States, Il Ciocco, Tuscany, Italy.

Adaptive automation provides a possible countermeasure to performance failure due to impaired operator functional states (OFS) in safety-critical systems. A model for matching adaptive automation to one component of OFS, mental workload, is described. The model is evaluated for two forms of adaptive automation, adaptive aiding and adaptive task allocation. A system for real-time, workload-matched adaptation based on physiological measures—heart rate variability and cerebral blood flow—is described and its use during multi-task performance evaluated. The adaptive system was associated with an overall enhancement of performance, a reduction in automation complacency, and a balancing of operator mental workload across variations in task load.

<u>Ongoing Studies</u>

1. Work on further development of the Viking platform is continuing. An initial study examining the effects of sub-task difficulty, sustained work, and automation on performance is currently being conducted.
2. A follow-up study to our previous experiment using the sensor-to-shooter simulation (Rovira, McGarry, & Parasuraman, 2002) is currently being programmed and will commence in September 2002. This investigation will extend that study to include examination of sustained workload and perceived stress.

Quarterly Progress Report
Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield
Army MURI, Subcontract to U. Central Florida

Raja Parasuraman
Catholic University of America

PERFORMANCE PERIOD: 7/1/02- 9/30/02

During this period, two studies were presented at the Annual Conference of the Human Factors and Ergonomics Society, Baltimore, MD.

Paper Abstracts

1. Rovira,E., McGarry, K., & Parasuraman, R. (2002). Effects of unreliable automation on decision making in command and control. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, Baltimore, MD, pp. 428-432.

The effectiveness of automated decision aids used by human operators in command and control systems may depend not only on automation reliability, but also on the type (stage) and level the automated support provides. Automation can be applied to information acquisition, information integration and analysis, decision choice selection, or action implementation (Parasuraman, Sheridan, & Wickens, 2000). The present study examined the effects of variations in the stage of automation support on performance in a "Sensor to Shooter" targeting simulation of command and control. Independent variables included the type and level of automation support (complete listing, priority listing, top choices, and recommendation of decision choice) and the reliability of the automation (60% and 80%). Dependent variables included accuracy and reaction time of target engagement decisions. Compared to manual performance, reliable automation did not affect the accuracy of target engagement decisions but did significantly reduce decision times. When the automation was unreliable, under the higher reliability condition (80%) there was a greater cost in accuracy performance for higher levels of automation aiding (priority listing, top choice, and recommendation) than at a lower level (complete listing). The results support the view that automation unreliability has a greater performance cost for decision automation than for information automation. This performance cost generalizes across a number of different forms of decision-aiding.

2. Rovira,E., Zinni, M., & Parasuraman, R. (2002). Effects of information and decision automation on multi-task performance. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, Baltimore, MD, pp. 327-331.

Automation purported to assist human operators may itself be an additional source of complexity and uncertainty. Because high reliability cannot always be assured, imperfect automation can add to uncertainty and thereby degrade performance. The

present study examined the relative benefits and costs of information and decision automation and investigated the effects of uncertainty resulting from automation unreliability during multiple task performance. Subjects were either provided with status information ("information" automation) or a recommendation for action ("decision" automation) for the system monitoring sub-task of the Multi-Attribute Task Battery (MAT). Two levels of automation reliability were compared. The detrimental effect of unreliable automation—a decrease in the detection rate of malfunctions—was greater for automation of higher reliability, a result consistent with previous findings of automation-related complacency. This effect of automation unreliability was also greater for decision than for information automation.

Ongoing Studies

1. An initial study examining the effects of sub-task difficulty, sustained work, and automation on performance on the Viking soldier platform is currently being conducted.
2. A follow-up study to our previous experiment using the sensor-to-shooter simulation (Rovira, McGarry, & Parasuraman, 2002) is currently being conducted. This investigation is extending that study to include examination of sustained workload and perceived stress.

Equipment Installation

` Under a separate grant, a Nicolet Pioneer TC8080 2-channel Transcranial Doppler Sonography System is currently being installed and tested in my laboratory. This system will be available full time for studies carried out under the MURI grant. We plan an initial study to assess its utility for real-time measurement of mental workload by monitoring cerebral blood flow in the left and right half brain of operators working on a command and control task.

The Effect of Physical Workload on Cognitive Performance

W. C. Harris, A. J. Walker & A. S. Leon
University of Minnesota

P. A. Hancock
University of Central Florida

ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Introduction

- Infantry equipment is becoming increasingly complex
- The infantry battlefield demands high levels of physical exertion
- Will physical exertion decrease cognitive performance and thereby limit soldier's performance during combat?

Objectives

- Assess the effect of current exercise level on cognitive performance
- Examine cumulative exercise effects on cognitive performance during a period of moderate exercise (5-mile march)

Method

- Subjects
- Cognitive assessment
- Procedures

Method

SUBJECTS

- 3 females and 7 males
- 19 to 38 years of age (Mean = 24.4)
- Military personnel
  - University of Minnesota R.O.T.C.
  - National Guard
  - Reserves
  - Active duty

COGNITIVE ASSESSMENT

- Automated Neuropsychological Assessment Metric (ANAM)
  - 6 sub-tests
- Test selection criteria
  - Minimize psychological demand overlap task
  - Sample critical information processing components

Sub-test Selection

- Information processing stages (Walter Reed PAB)
  - Perceptual input, detection & identification
  - Central processing
    - Short- and long-term memory, associative memory
    - Information integration/manipulation
      - Linguistic/symbolic
      - Spatial

Sub-tests

- Simple reaction time
- Coding
- Match to sample
- Memory (2 and 6 item)
- Spatial processing
- Logical reasoning

Procedures

Cognitive Task Training – to stable baseline

Treadmill

- Intensity
  - Low intensity (30% VO2 Max)
  - Moderate intensity (60% VO2 Max)
  - High intensity (90% VO2 Max)

- Five - mile march (75% VO2 Max)

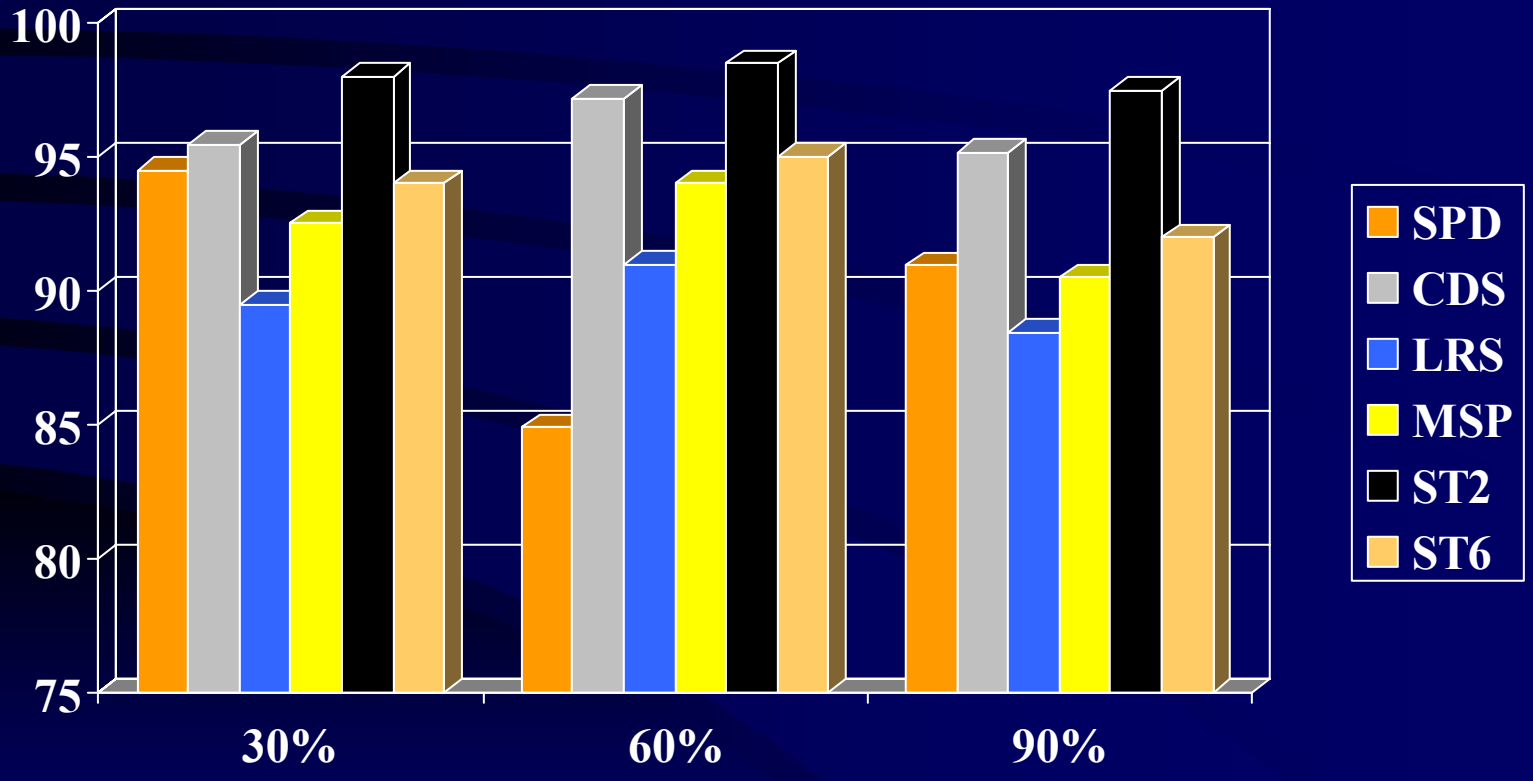
Results

- Exercise level (acute effect)
  - Accuracy
  - Response Time
- Five-mile march (sustained effort)
  - Accuracy
  - Response Time

Accuracy During Three Exercise Levels

		LEVEL		Signif.
	30%	60%	90%	P =
CDS	95.46	97.19	95.15	0.19
SPD	94.50	84.87	90.95	0.03
LRS	89.50	91.00	88.45	0.41
MSP	92.50	94.00	90.50	0.67
ST2	98.00	98.50	97.50	0.56
ST6	94.00	95.00	92.00	0.64

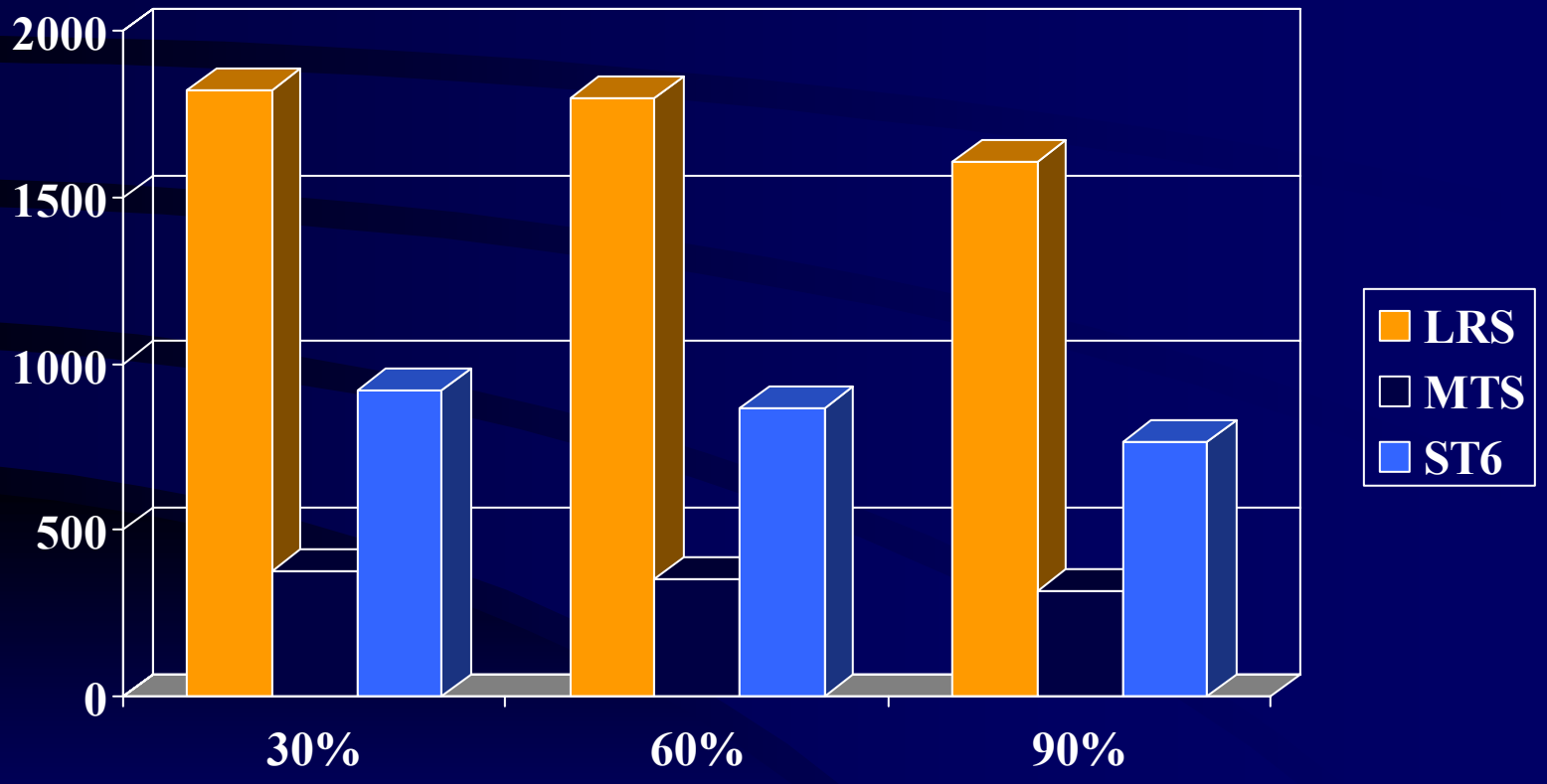
Accuracy



Cognitive Response Time at Three Exercise Levels

	30%	60%	90%	P =
Task				
SRT	426.9	445.2	421.0	0.16
CDS	1068.1	1086.55	1051.4	0.31
SPD	1731.1	1609.3	1561.12	0.13
LRS	1820.65	1797.05	1604.05	0.01
MTS	376.37	354.59	317.8	0.07
ST2	644.6	645.85	610.9	0.14
ST6	916.9	867.45	761.5	0.05

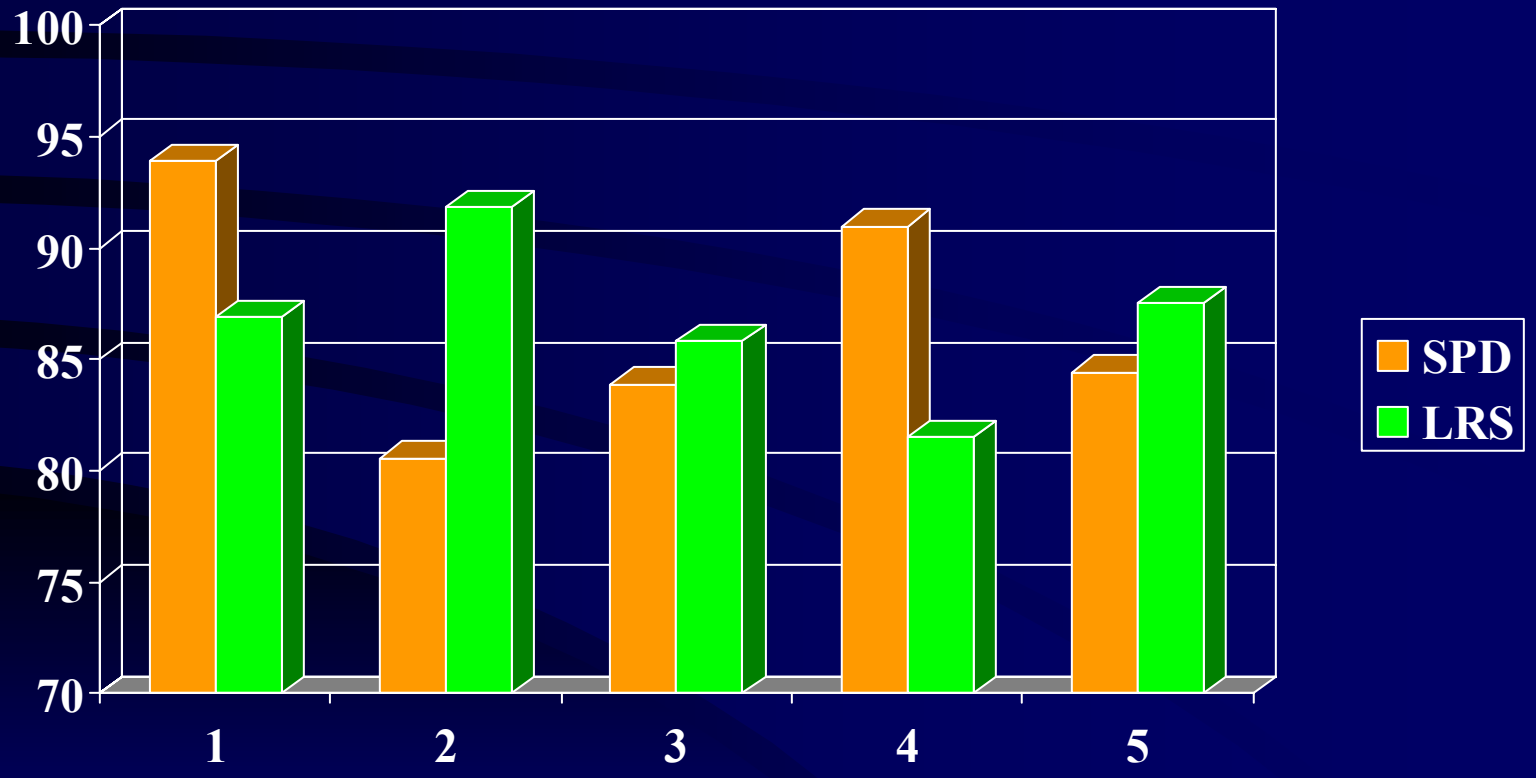
Response Time at Three Exercise Levels



Accuracy during 5-Mile March

			TEST			
	1	2	3	4	5	P =
CDS	96.36	97.22	94.80	94.80	96.88	0.13
SPD	93.89	80.56	83.89	90.97	84.44	0.001
LRS	86.88	91.81	85.82	81.51	87.50	0.06
MSP	93.57	92.14	92.86	90.71	89.29	0.61
ST2	95.71	95.71	96.43	96.43	96.43	0.99
ST6	93.57	93.57	94.29	92.86	87.86	0.48

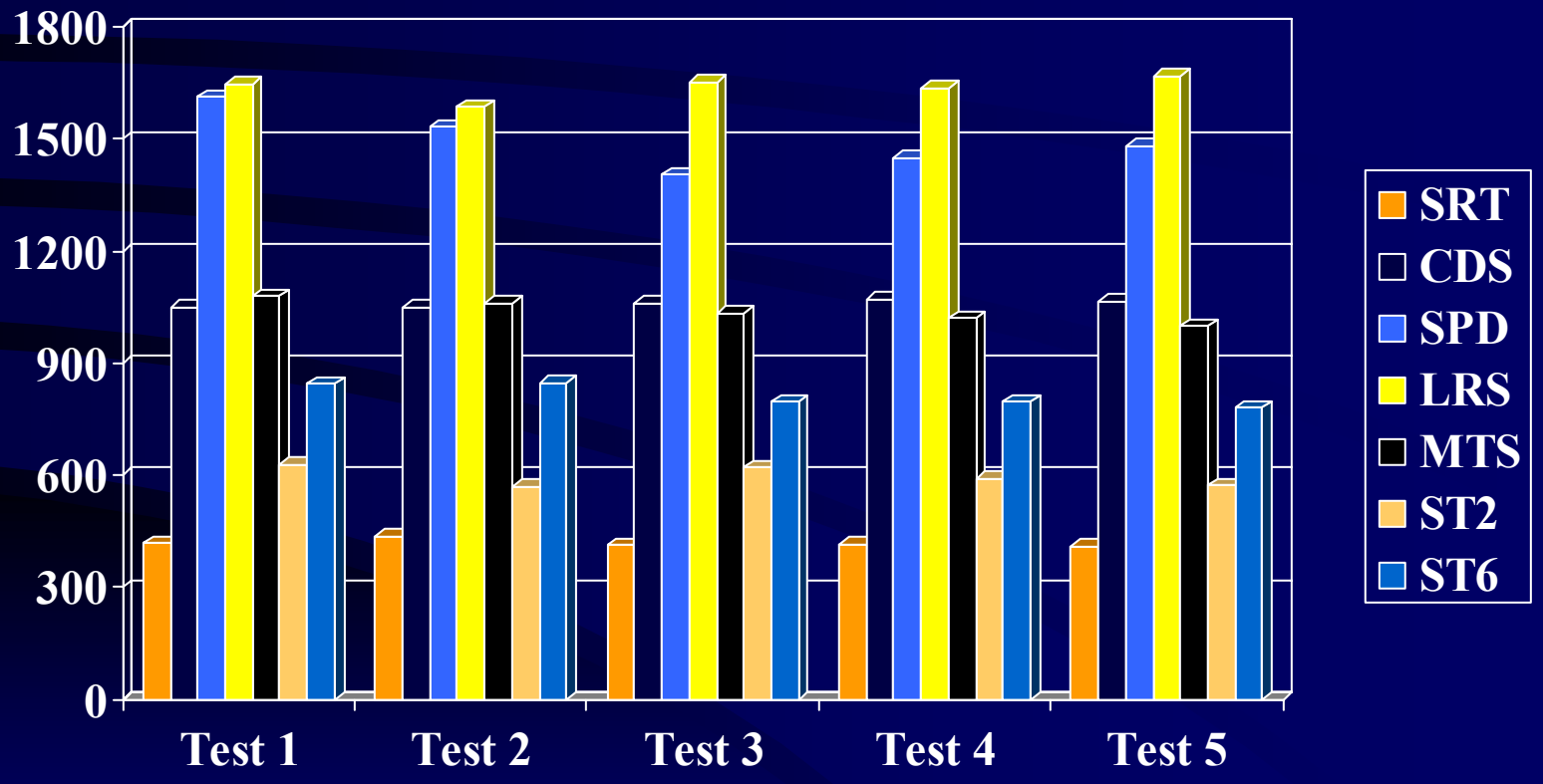
Accuracy



Response Time during March

	Test 1	Test 2	Test 3	Test 4	Test 5	P =
Task						
SRT	418.06	435.56	413.33	415.0	408.67	0.46
CDS	1049.7	1050.2	1061.2	1071.3	1065.2	0.99
SPD	1611.0	1531.9	1405.1	1450.3	1481.5	0.41
LRS	1645.9	1585.1	1653.3	1635.2	1669.3	0.90
MTS	1078.3	1059.0	1033.1	1019.1	999.0	0.23
ST2	629.43	570.36	620.36	590.0	571.29	0.17
ST6	843.57	847.0	797.36	797.07	781.21	0.27

Response Time during March



Summary

1. Tendency for faster, less accurate responding at high exercise levels. Change varied with task.
2. Accuracy changes were small, and not consistent over tasks. Only spatial accuracy decreased significantly.
3. Response time improved during intense exercise level (grammatical reasoning and memory)
4. Accuracy was stable during march except for SPD decrease
5. Response time was stable during 5-mile march

Conclusions

- Minor cognitive performance changes were observed at the exercise levels examined
- Spatial processing accuracy was most susceptible to the effects of exercise
- Response speed increased during high exercise periods and complex tasks seem to be affected most

Questions?





Infantry Cognitive Changes During Simulated Combat





Wayne C. Harris, Ph.D.

Tammy L. Grady, 1LT, WIARNG

Human Factors Research Laboratory
University of Minnesota



ACKNOWLEDGEMENT

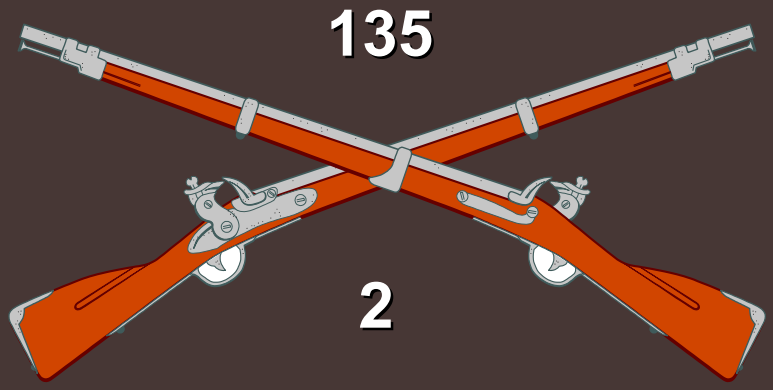
The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Inquiries

- Dr. John Chattoe, University of Toronto
- Dr. Susan Hill, Human Research and Engineering Group
- Dr. Richard Christ, Army Research Institute

Ft. Chaffee Study

2nd Battalion 135th Infantry (Air Assault)



Introduction



- Increasing system complexity increases the amount and complexity of the information soldiers must process
- Stress impairs information processing
- Soldier's information processing during combat needs to be known in order to design equipment for the combat environment

Project Objective

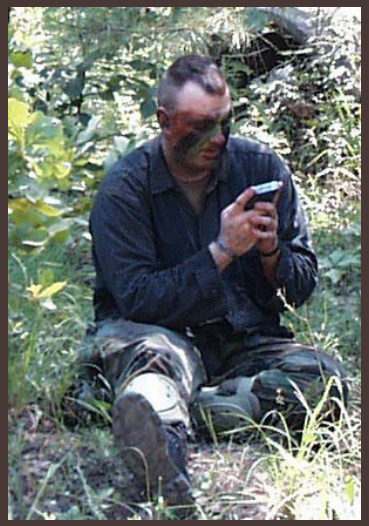
- Determine the effect of the infantry operational environment on specific information processing components





Information processing components







Study Outline

- Identify a setting representative of the infantry operational environment
- Identify potential participants who are representative of combat personnel



- Train participants to perform the assessment battery
- Determine a non-combat operational environment performance level
- Assess performance during (or in close proximity to) a period of simulated combat

Data Collection

Days 1 & 2	**Barracks**	**Training & Baseline Assessment**
Days 3-8	**Field Exercise**	**Day 5 Field Assessment 1** **Day 7 Field Assessment 2**
Day 9	**Barracks**	**Day 9 Assessment**
Days 10-12	**Barracks/Field**	
Day 12	**Assault**	**Field Assessment 3**

Cognitive Processes Measured

- Reaction time
- Spatial working memory



- "Sternberg memory"
- Abstract reasoning
- Mathematical processing
- Spatial processing

ARES Cognitive Task Examples

REACTION TIME

*

Subject's task: Press key when the star appears.

ABSTRACT REASONING

& is first

& #

Subject's task: Indicate whether statement is true or false.

MATHEMATICAL REASONING

5 + 7 - 4

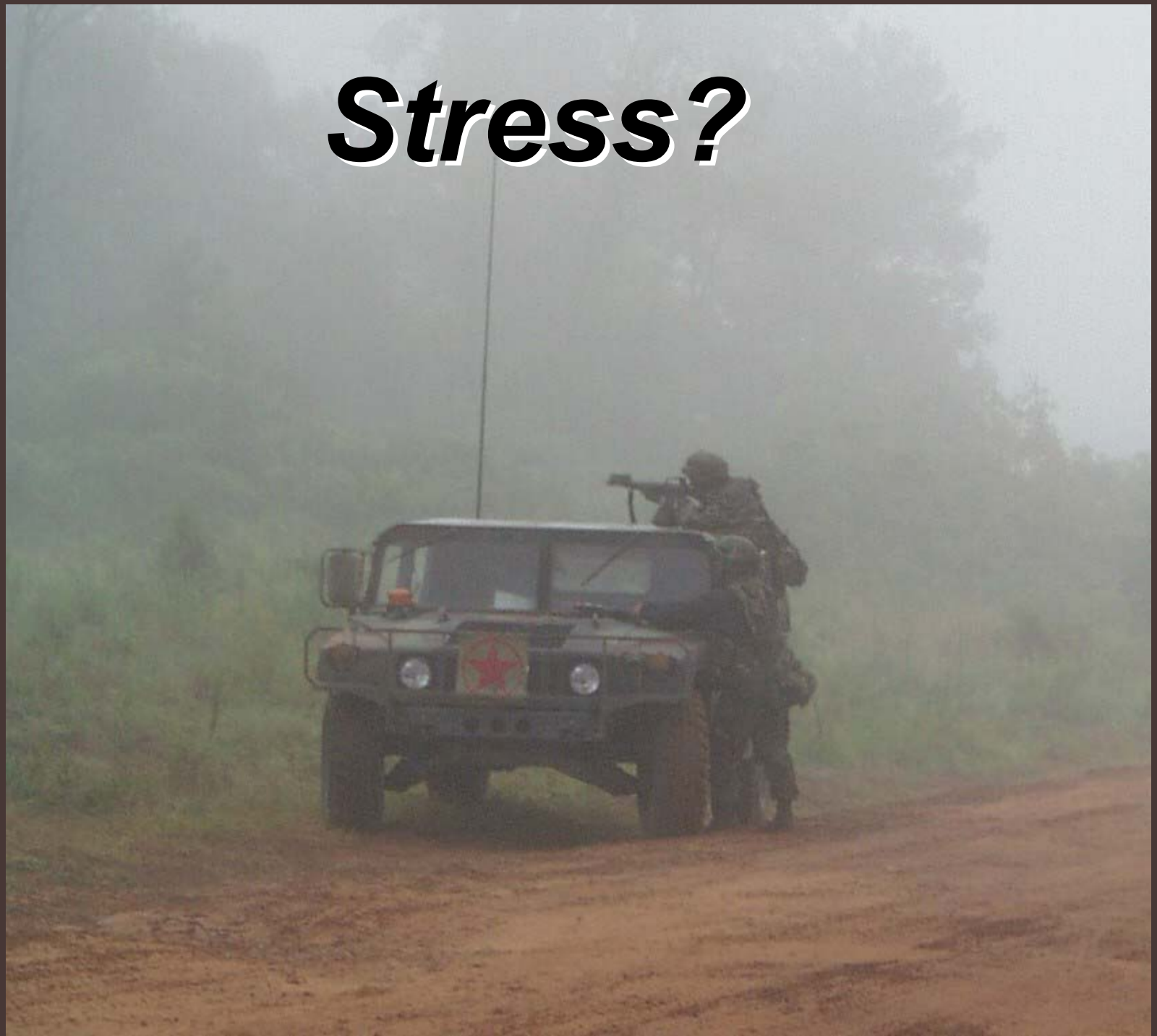
Subject's task: Indicate whether result is greater or less than 5.

The Infantry Field Environment

- Normal operational demands
  - Discomfort (heat, cold, noise . . .)
  - Physical exertion
  - Sleep (time & quality)
  - Mental workload, vigilance
  - Danger (snakes, falls, insects . . .)
- High operational demands
  - Immediate threat of severe injury or injury



Stress?



Data Analysis

- Battalion changes
- Alpha vs. Charlie Company changes
- Near casualty changes



Major Finding

- Soldiers tolerate field conditions with minimal disruption of cognitive performance

- High demand events are associated with cognitive performance changes

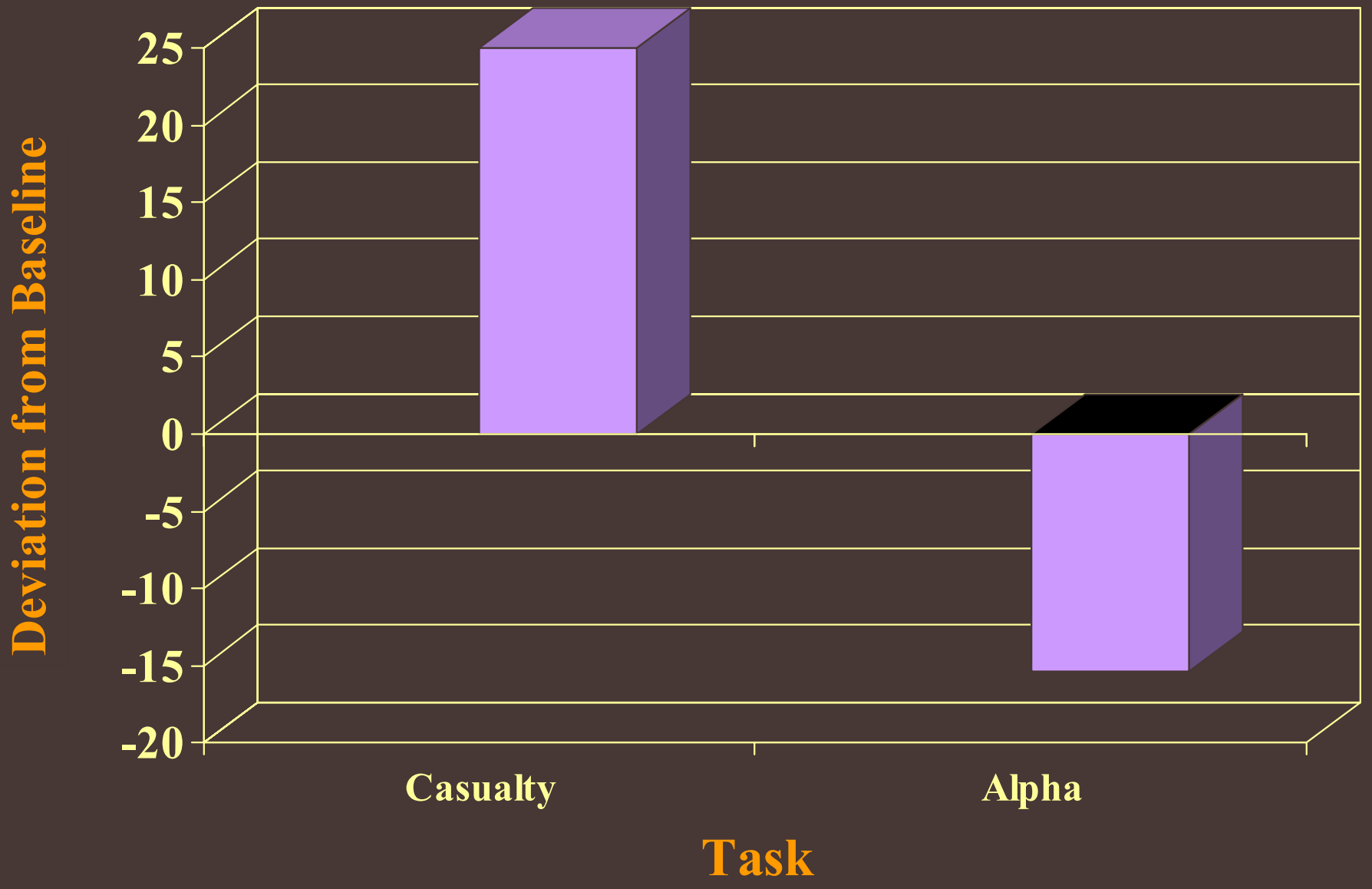


Near Heat Casualty

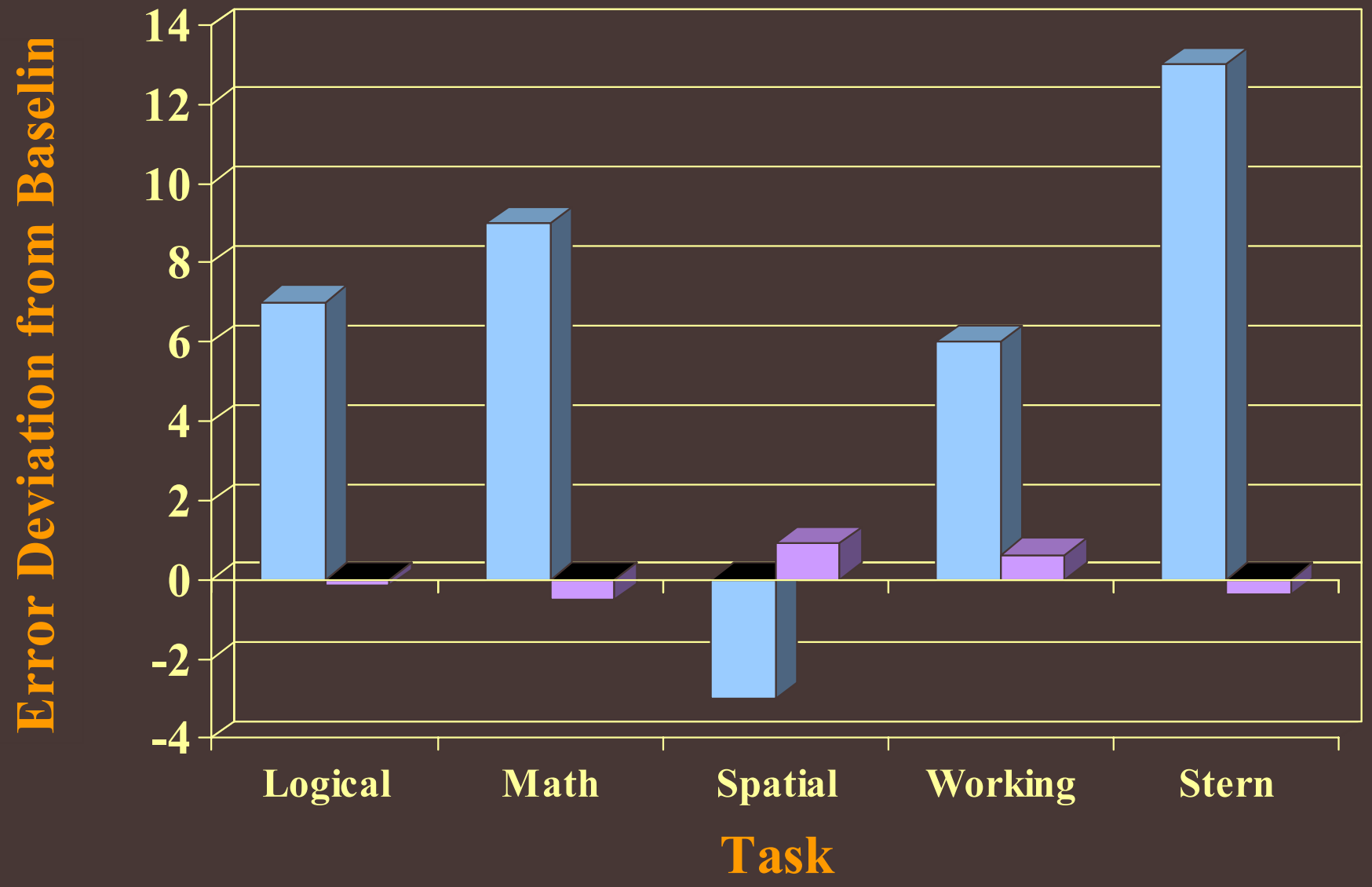
- Simple reaction time
- Accuracy
- Response time



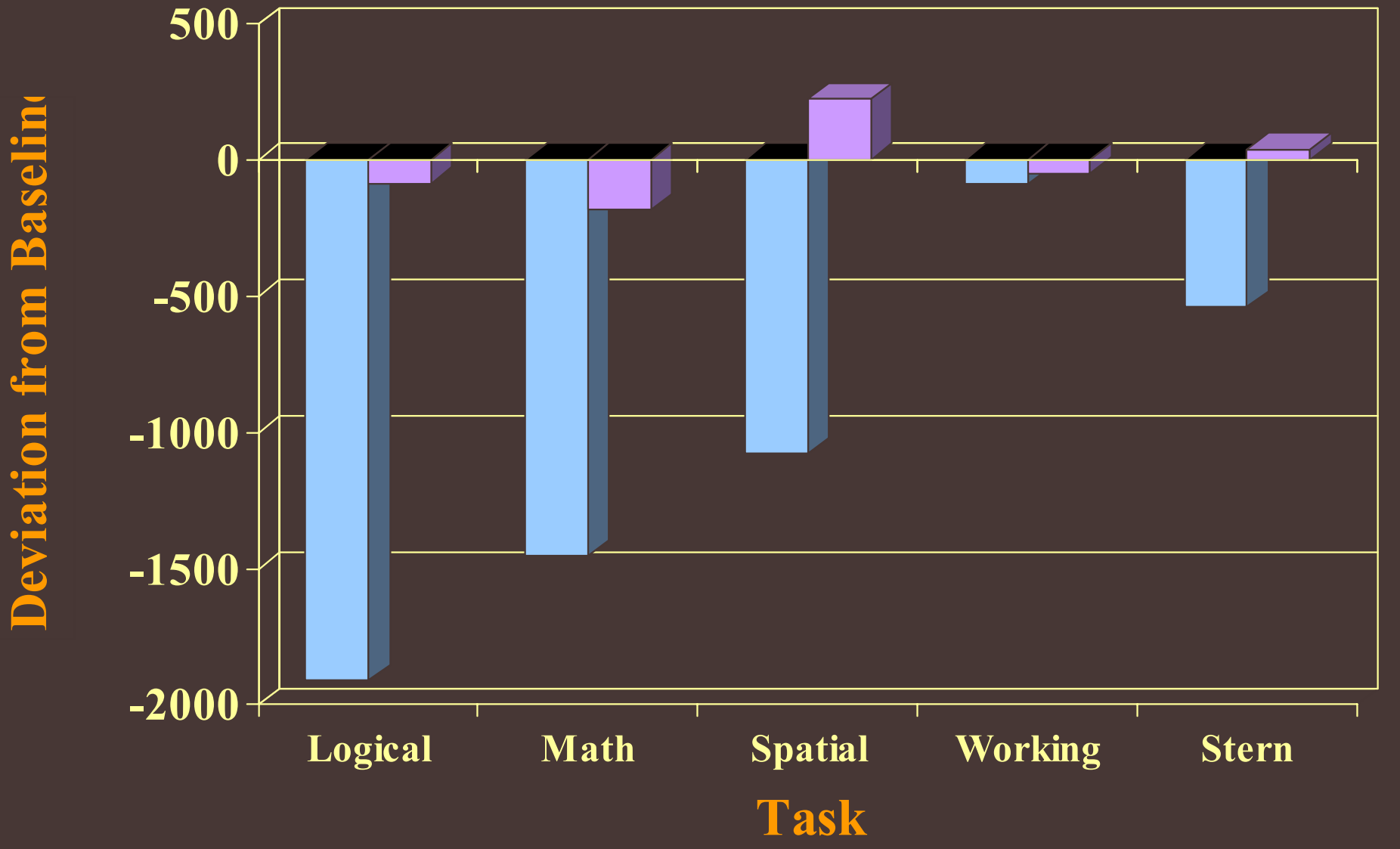
Simple Reaction Time



Errors


Casualty
Alpha
Error Deviation from Baselin
14
12
10
8
6
4
2
0
-2
-4
Logical
Math
Spatial
Working
Stern
Task

Response Time



Lessons Learned

- Training provides an excellent setting to examine the effects of normal operational demands
- Incidents provide data representative of high operational demands
- Demand level varies between units and between individuals within units (you need to be there)
- High threat conditions disrupt cognitive performance
- Cognitive tasks do not appear to be affected equally







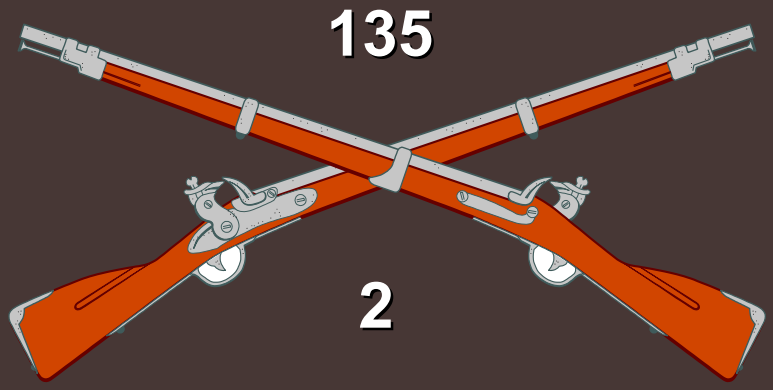
Questions?





Ft. Chaffee Study

2nd Battalion 135th Infantry (Air Assault)



The VIKING Experimental Environment

A Multi-task Infantry Information/communications Environment

- **W. C. Harris**
  Human Factors Research Laboratory
  University of Minnesota
- **R. Parasuraman & M. Zinni**
  Cognitive Science Laboratory
  Catholic University of America
- **P. A. Hancock**
  MURI Research Laboratory
  University of Central Florida
- **S. C. Harris**
  Human Research and Engineering Directorate
  U.S. Army Research Laboratory



ACKNOWLEDGEMENT

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Tailoring Technology to the Combat Environment

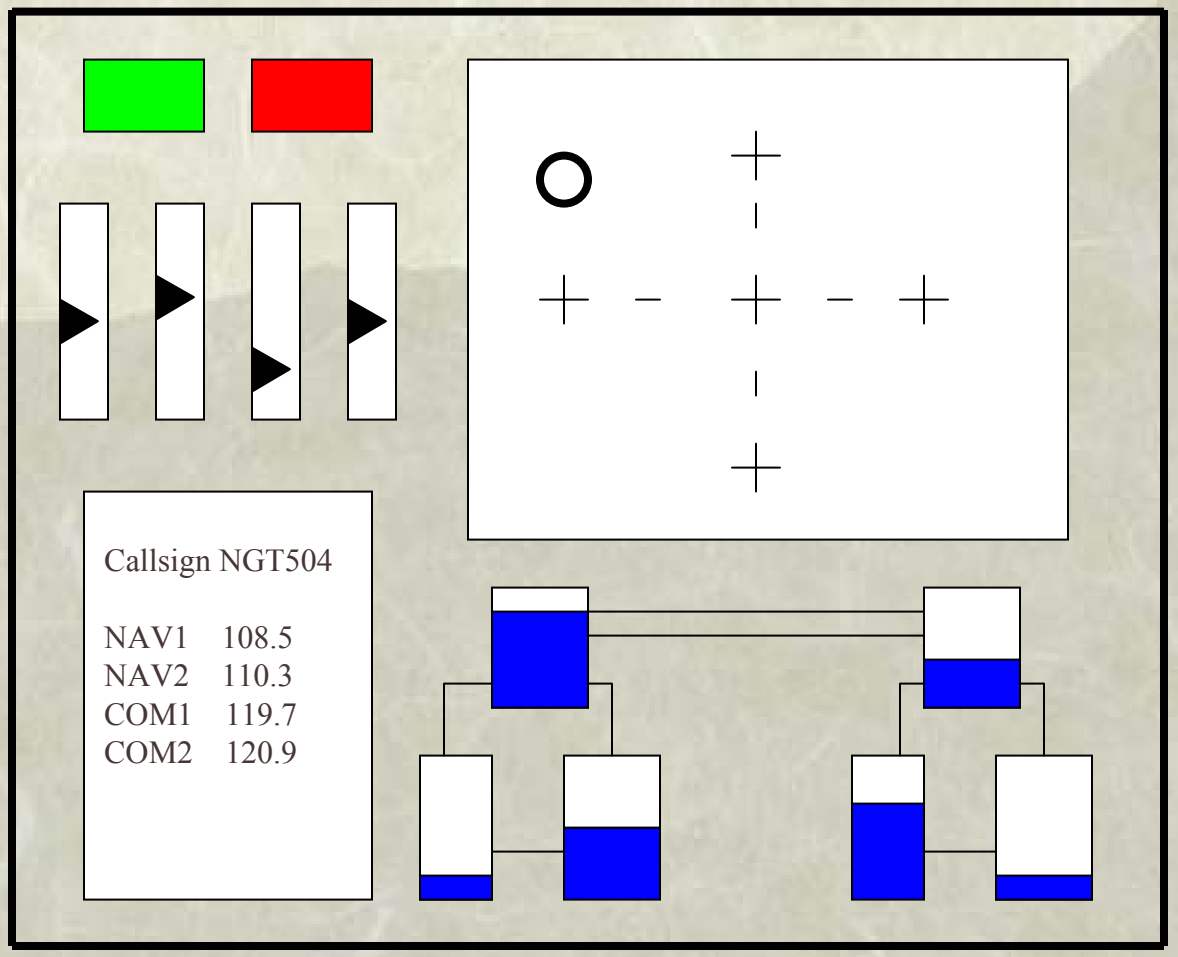
More than an engineering problem



The VIKING Experimental Environment

- Represents a limited sub-set of infantry tasks – no shooting or moving
- The environment provides a platform to examine issues such as the effect of stress, workload, fatigue, and automation
- Software is a modification of the NASA Multi-attribute Task Battery (Comstock and Arnegard, 1992)

MAT



NASA MATB

Environmental model: Aviation

Task class	Cognitive demand (Parasuraman)
Tracking	Perceptual-motor
Monitoring	Perceptual-cognitive
Resources management	Cognitive-strategic
Communication	Auditory input

MURI VIKING

- Environmental model: Infantry soldier
- Multiple tasks
- Tasks are performed simultaneously
- Cognitive demands of tasks are related to:
  - Operational systems
  - Cognitive processing stage
- Cognitive demands of events are determined by pre-session briefing
- Events are under experimenter control
- Automatic performance data collection

VIKING

Event and Response Options

- **Events**
  - **Digital & analog (monitoring/resource management)**
  - **Simple to complex**
- **Response Selection**
  - **Detection**
  - **Detection & data evaluation**
    - Simple to complex
    - Numerical/text/spatial
- **Information acquisition**

PROCESSING STAGES

- Perceptual input, detection & identification
- Central processing
  - Memory
  - Information integration/manipulation (symbolic)
  - Information integration/manipulation (spatial)
- Output/response execution
- Divided attention

Perceptual Input, Detection, & Identification

Cognitive demand	Viking task
Stimulus detection & identification	Warnings & alerts Link maintenance
Status monitoring	Platoon spacing
Active data acquisition	Hostile unit detection

Central Processing

Cognitive demand	Viking task
Memory	Radio frequency change Alert messages Track identification Hostile unit detection
Symbolic processing	Track identification System configuration (Platoon spacing)
Spatial processing	Track identification

Response Execution

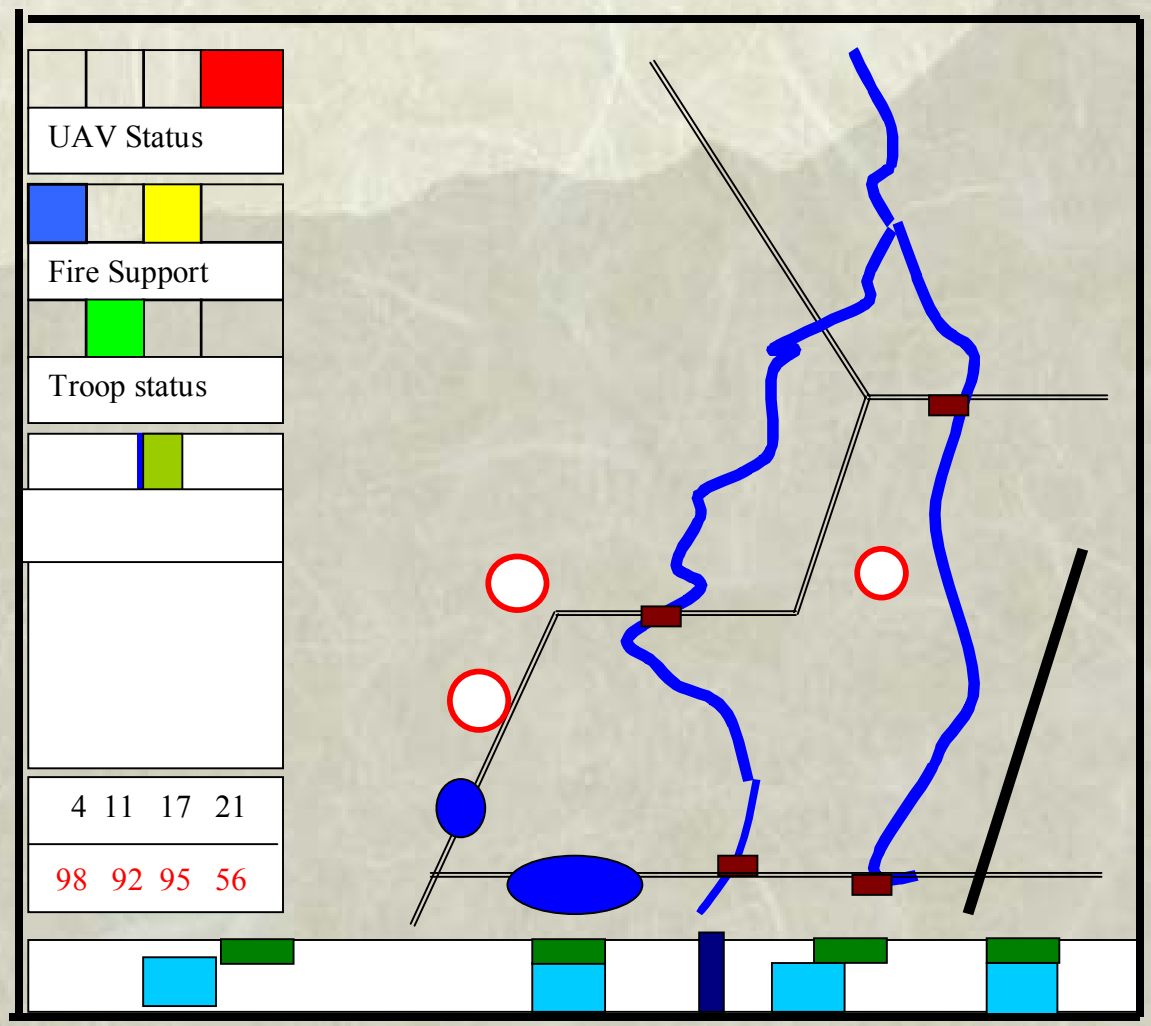
Cognitive demand	Viking task
Motor response	All tasks require keyboard responses

Divided Attention

Cognitive demand	Viking task
Divided attention	Multi-task environment with event control

The VIKING Interface

- Status lights
- Alert/warning information area
- Radio frequency selection and track data area
- Link performance data area
- Platoon unit location display
- Tactical plot

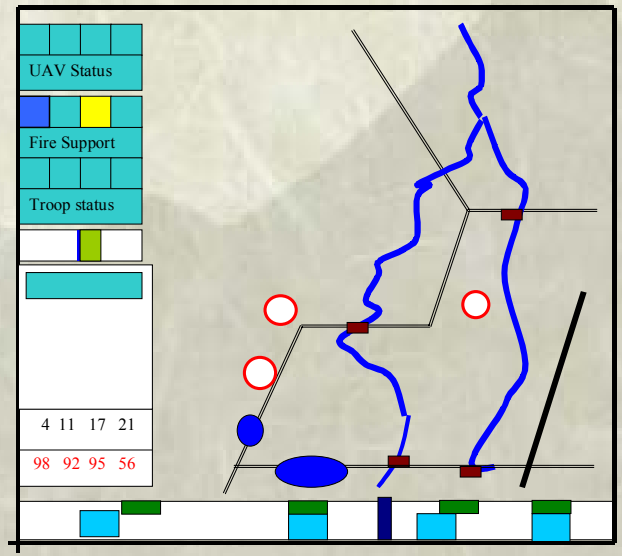

UAV Status
Fire Support
Troop status
4 11 17 21
98 92 95 56

Tasks

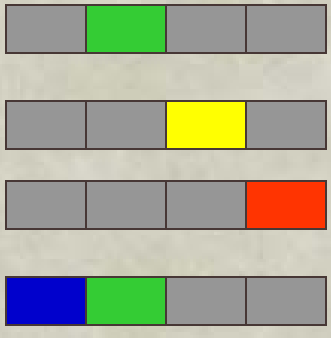
- Scenario based
- Detect & respond to lights
- Change radio frequency on command
- Detect link alert and evaluate numerical data
- Maintain platoon unit spacing and schedule
- Classify aircraft from data profile
- Elicit information to detect and destroy enemy units

Alerts & Warnings

- Three light displays



- Light display states
  - Normal
  - Alert
  - Warning
  - System active

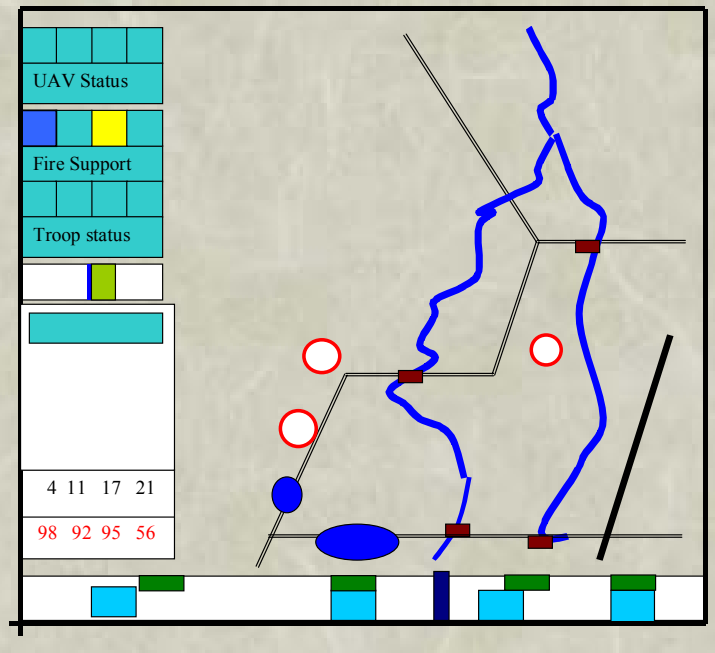


Responses to Alerts & Warnings

Warning F1, F2, & F3 (top/middle/bottom light)

Alert Press 1, 2, & 3 (top/middle/bottom light)

Response elicits the cause of an alert or warning

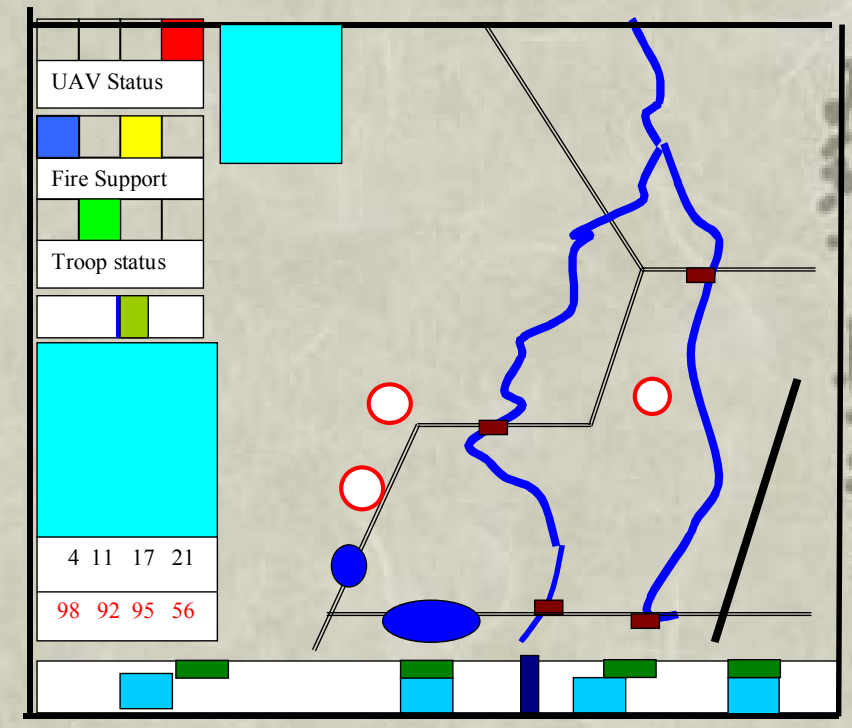


Radio Frequency Changes

B3	137.1
E2	132.6
E3	121.5
G5	141.2
I 3	145.8

Bravo 3

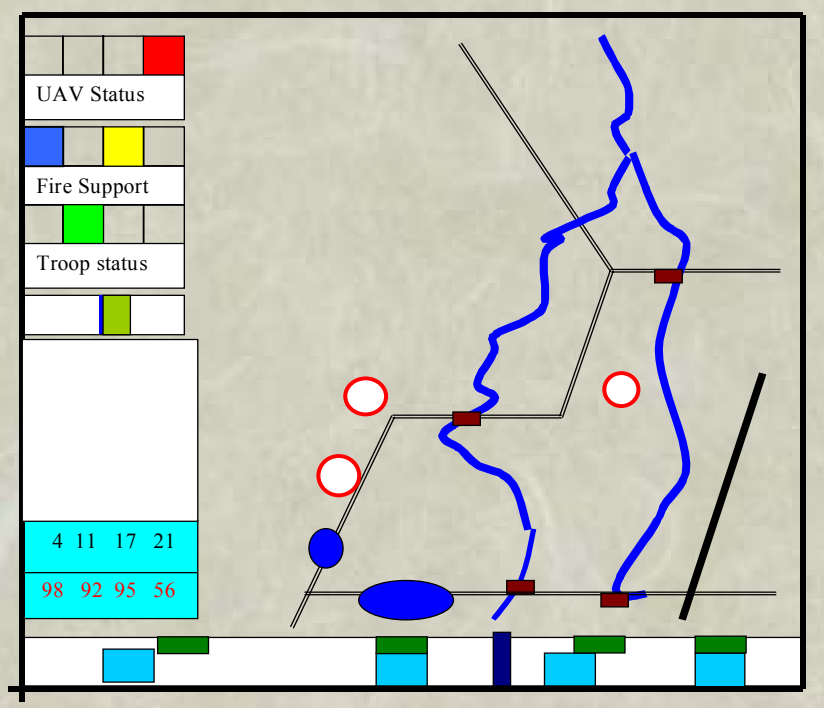
121.5
132.6
137.1
141.2
145.6
Highlight
& select



Links

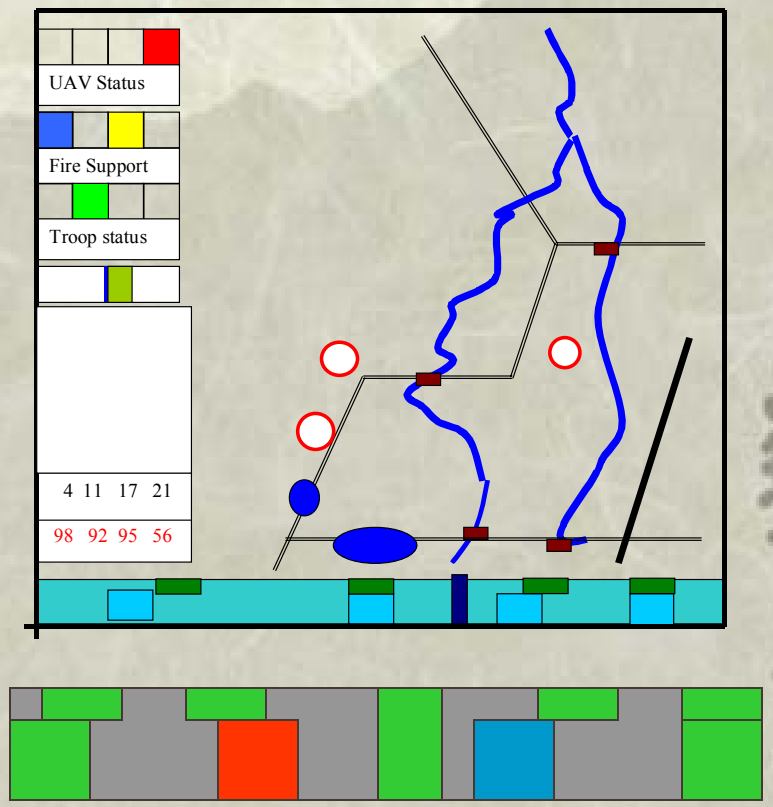
- Is any value less than 90?
- Yes – Indicate failed link by pressing F5 to F8 (left link to right link)
- No – Indicate false alarm by pressing F9

4	11	17	21
98	91	94	43



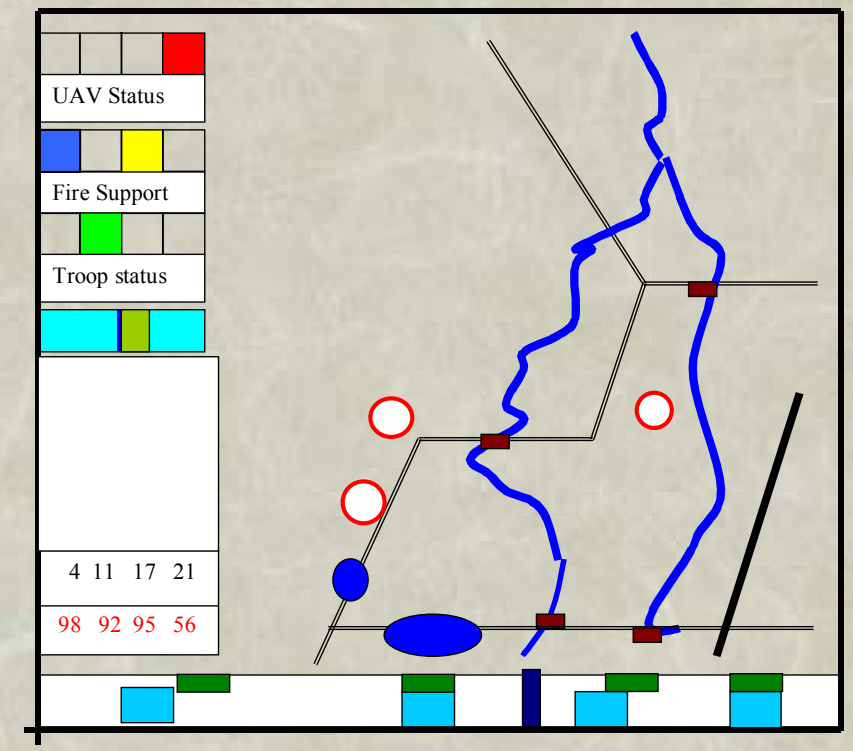
Platoon Spacing

- Keep units at spacing markers
- Unit color indicates movement status
- Start/stop units with keys Z, X, C, V & M
- B starts/stops all units



Pace Monitor

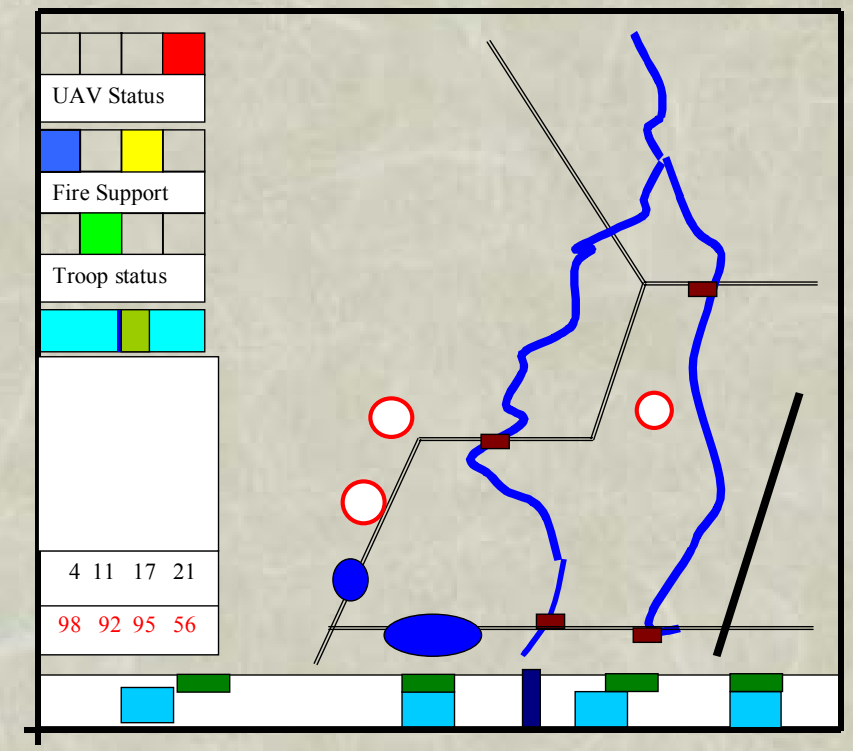
- Time/distance display begins with line in the middle of the bar (On schedule) 



Pace Monitor

- If the platoon is ahead of schedule, a green bar moves to the right

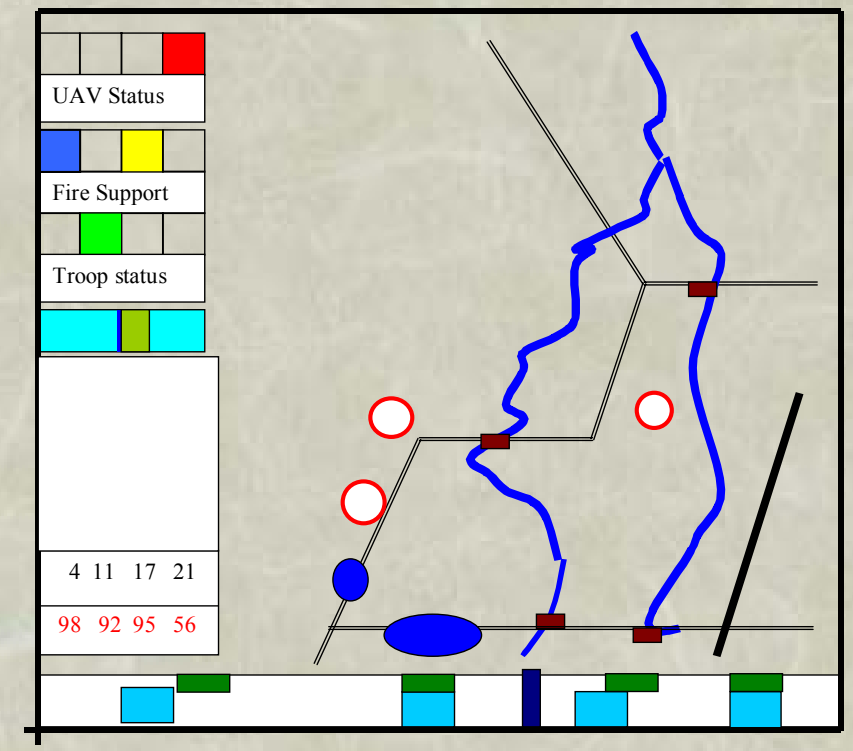




Pace Monitor

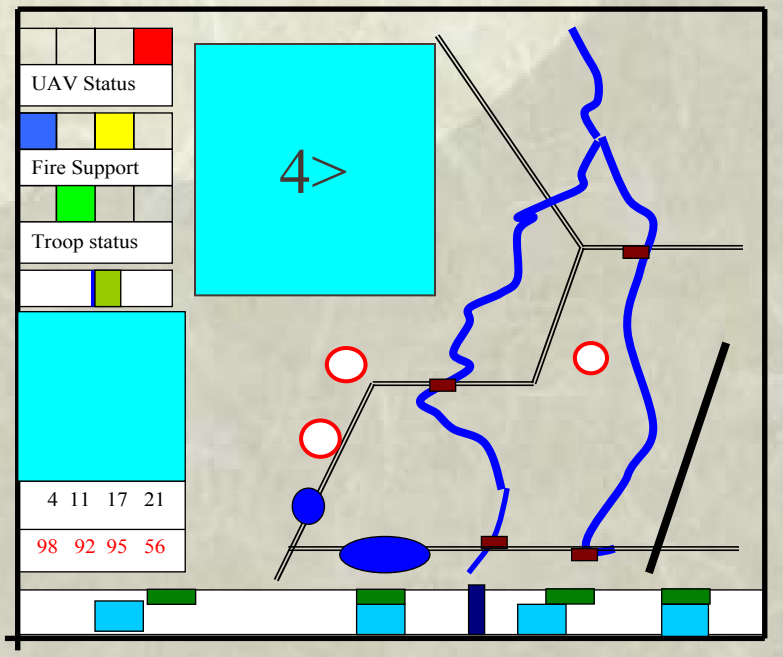
- If the platoon is behind schedule, a red bar moves to the left





Aircraft Identification

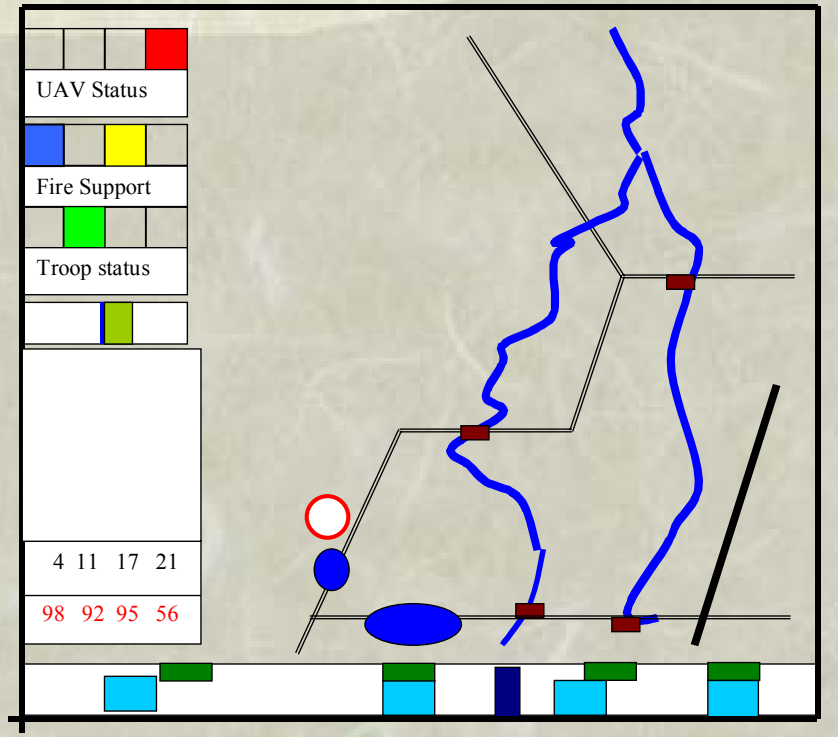
- Track appears on Tactical plot
- Data appears in message window
- Response
  - Friend (a)
  - Hostile (d)
  - Cannot classify (s)



Data	F	H
IFF	80	25
EW	20	25
Pos	10	50
V/A	45	50

Detect & Destroy Enemy

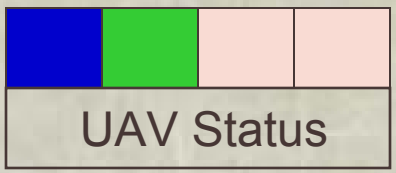
- Enemy unit appears automatically when platoon “makes contact”

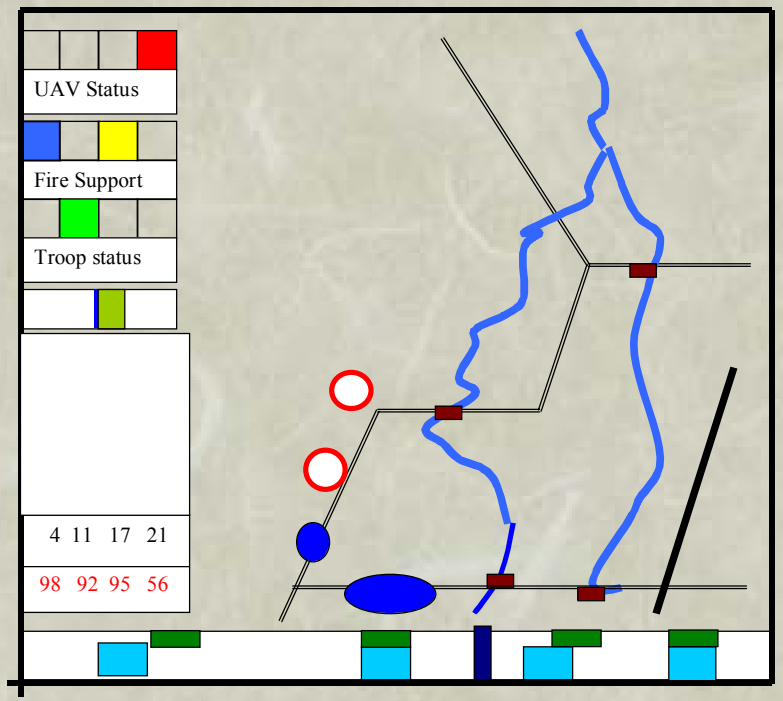


Enemy unit appears on map as:

Detect & Destroy Enemy

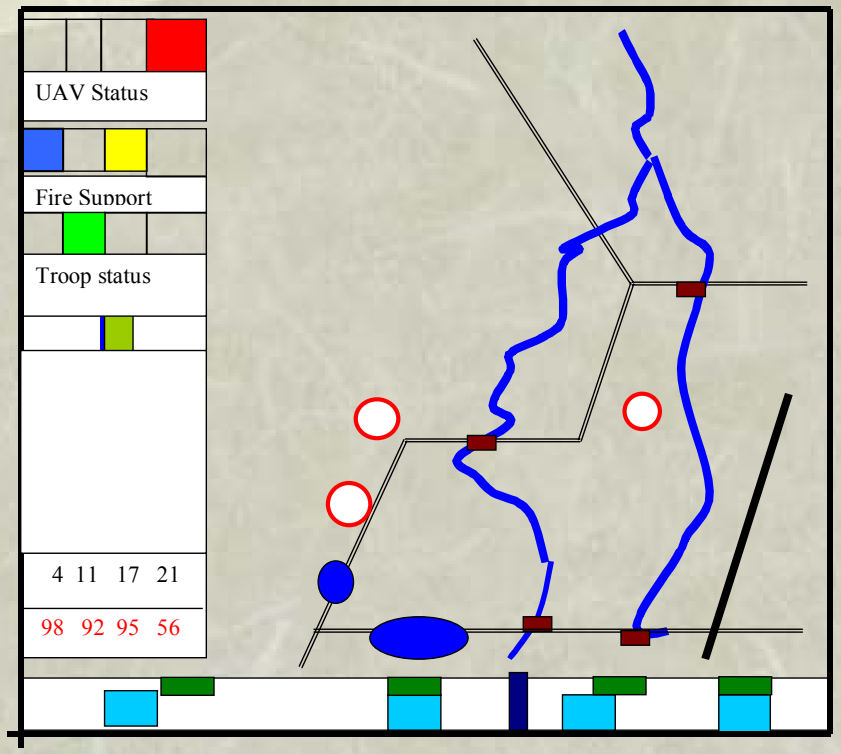
- Call for UAV (0) detects hostile unit before contact





Detect & Destroy Enemy

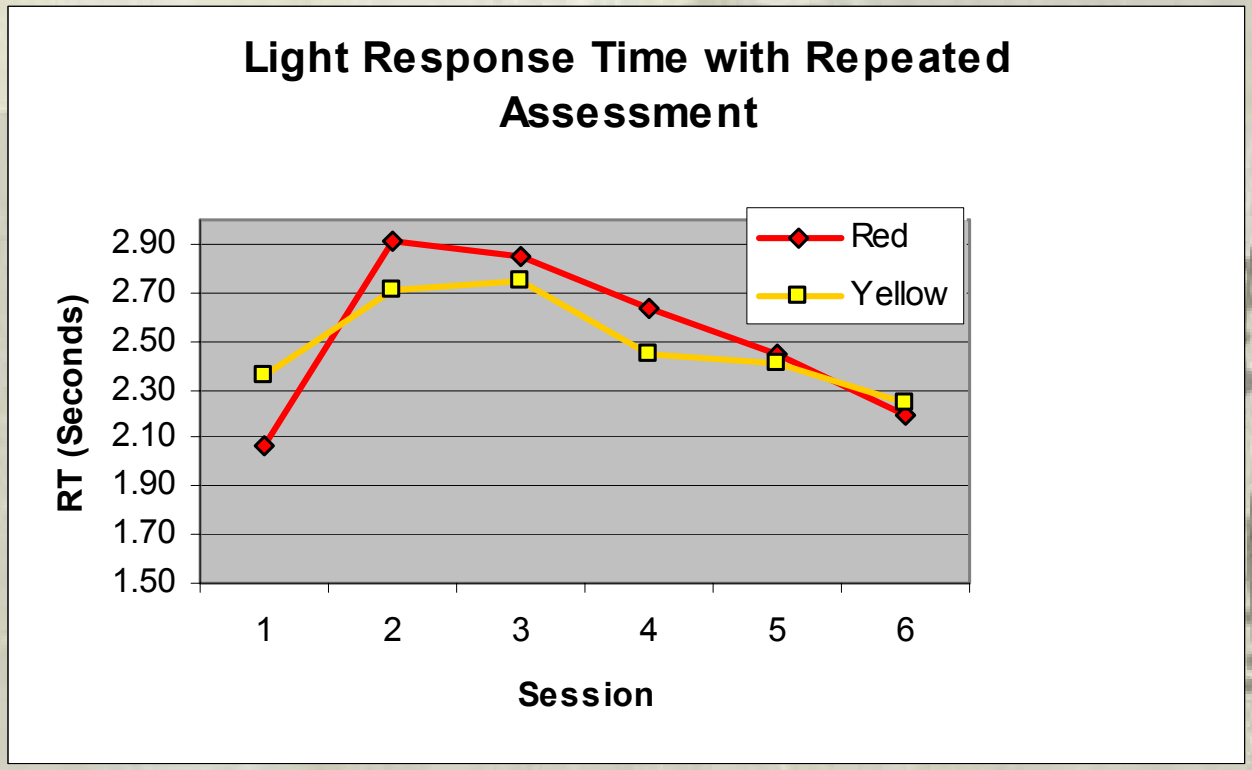
- Target (/) and destroy (p)



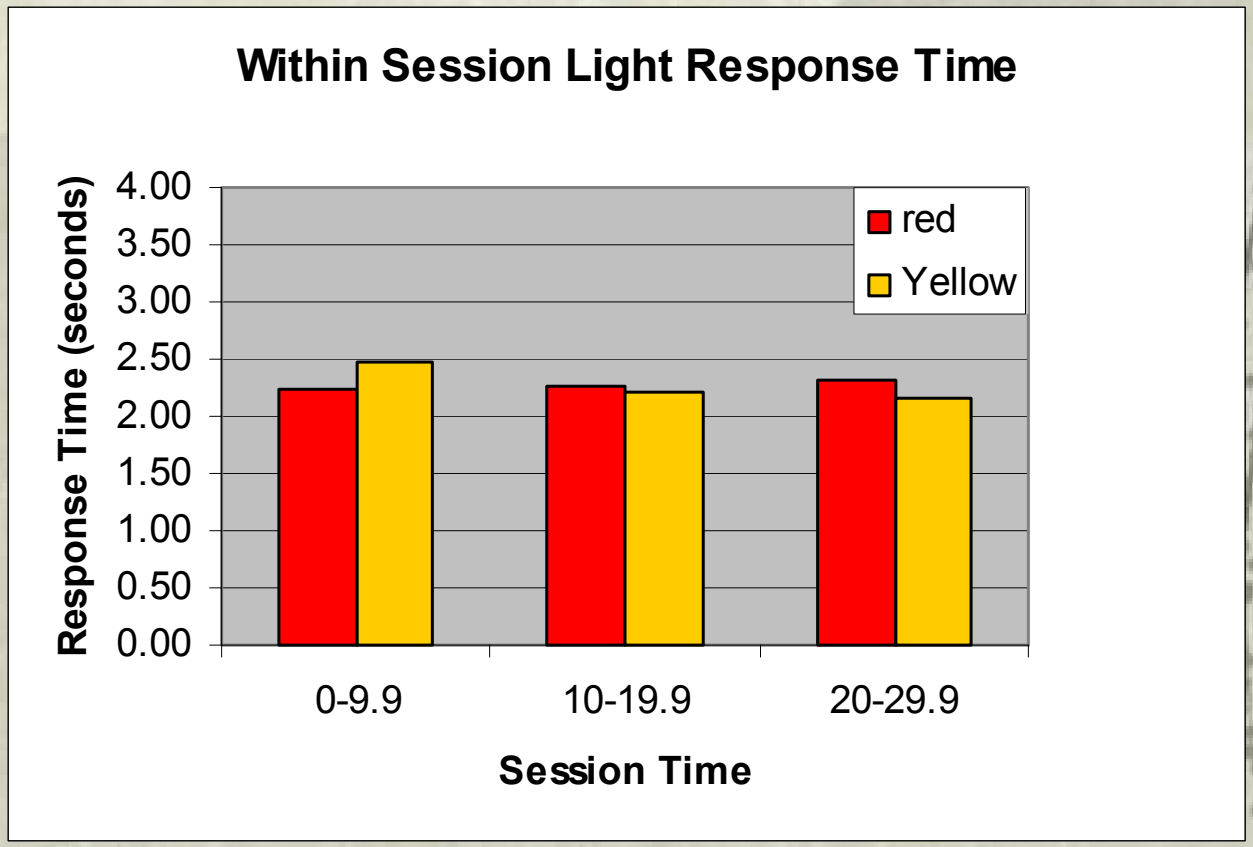
Summary

- The Viking Experimental Environment provides:
  - A multi-task environment
  - Tasks with different cognitive demands
- Tasks are relevant to infantry personnel

Baseline Data



Baseline Data



Viking Training Program



Situation

- You are the leader of the Third Platoon of Alpha Company

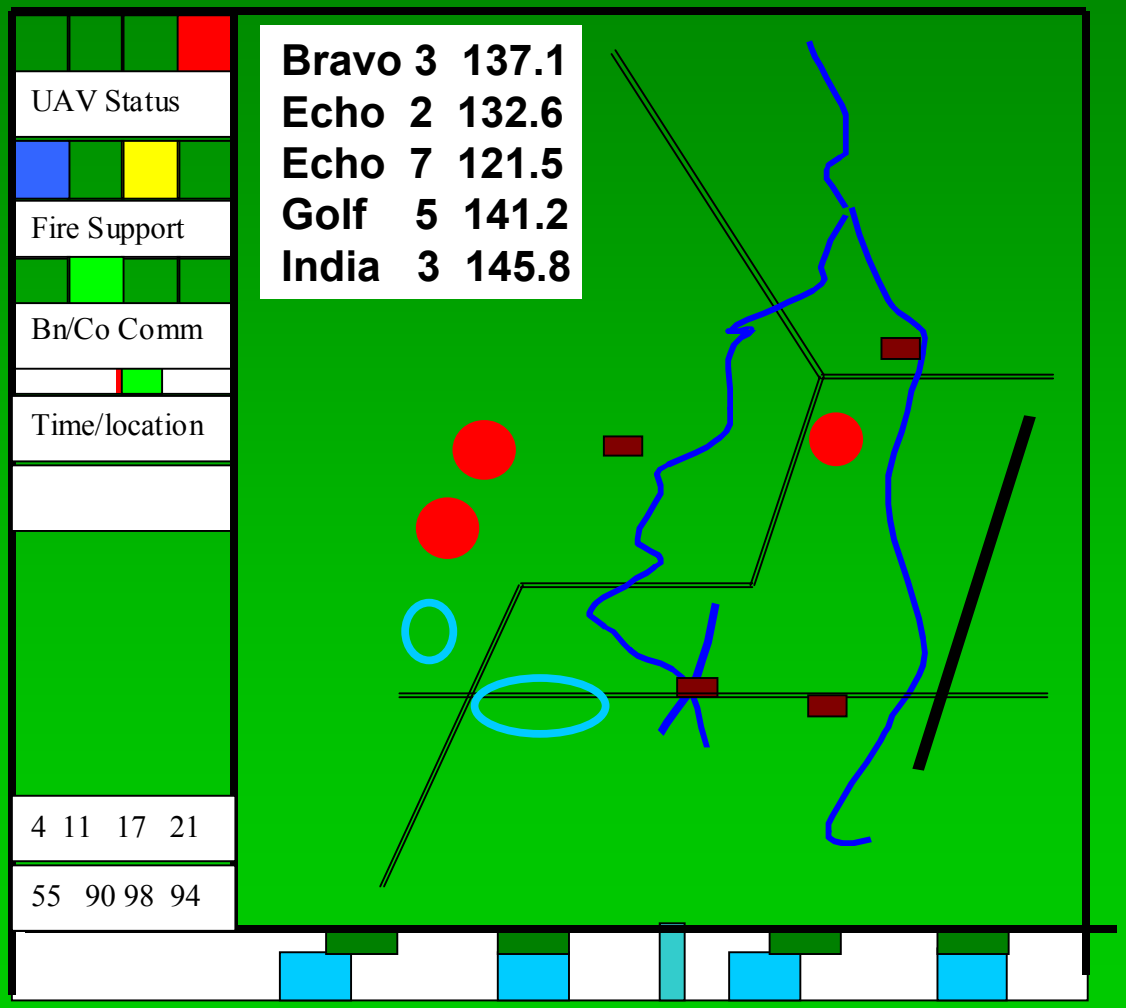
Mission

- Ensure the safe and timely arrival of your platoon to rally point for follow-on mission

Execution

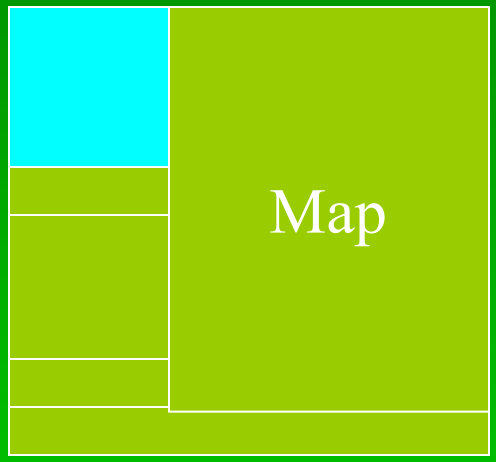
- Detect and Respond to Lights
- Maintain links
- Change radio frequency on command
- Maintain platoon spacing
  - And issue orders to maintain proper spacing
- Classify aircraft from data profile
- Elicit information to detect and destroy enemy units
- Reach rally point

What your screen will look like ...

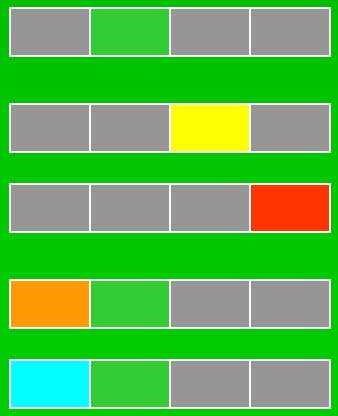


Detect and Respond to Lights Alerts & Warnings

- Three light displays (in upper left corner of display)
  1. UAV Status (Air Reconnaissance availability)
  2. Fire Support
  3. Bn/Co Communication

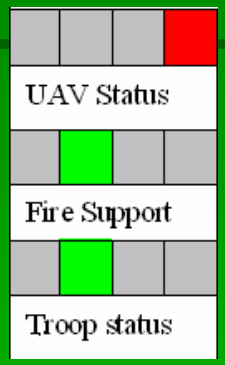


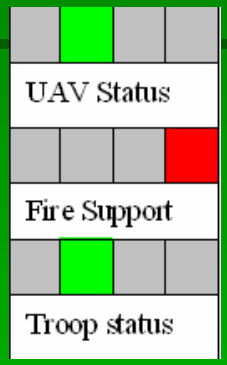
- Light display states
  - Normal (steady state) green
  - Alert yellow
  - Warning red
  - System Engagement orange
  - System active blue

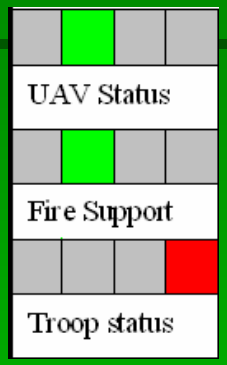


Detect and Respond to Lights Alerts & Warnings (Cont.)

- If







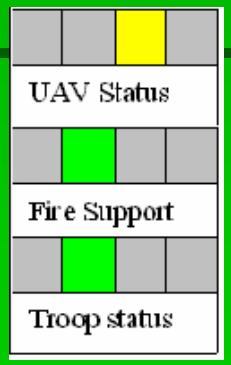
push

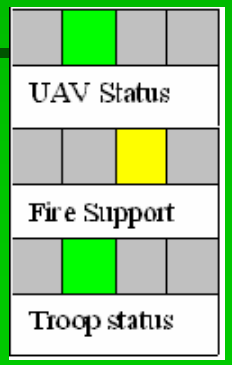


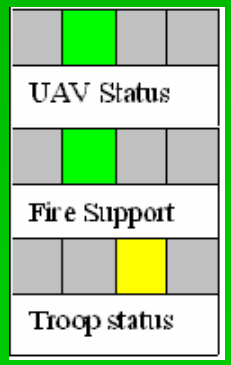




- If







push



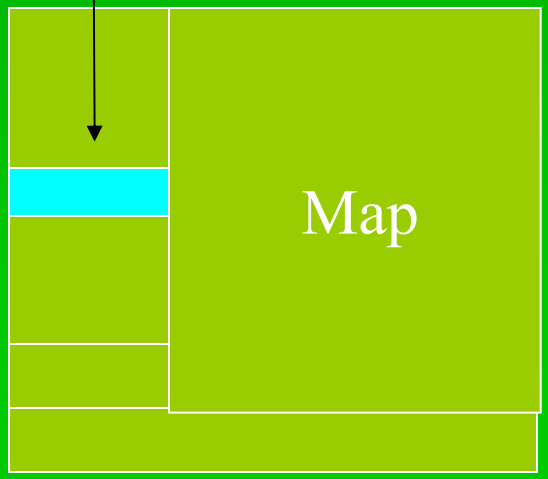




Detect and Respond to Lights Alerts & Warnings (Cont.)

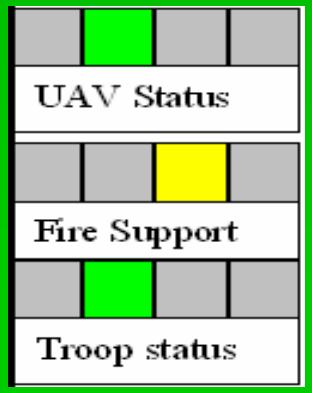
- Note: After you respond correctly to the light …
  - a message will appear instantly (in the message area)
  - the light will NOT disappear instantly
  - The light and message will disappear automatically after a few seconds

(you only need to press the correct key ONCE)



Detect and Respond to Lights Alerts & Warnings (cont.)

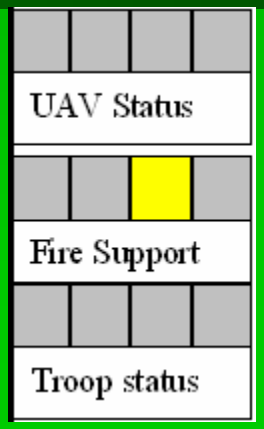
- Let's try it ….
- Note: other than the ⇦ & ⇨ (left & right arrow) keys, please do not actually press any keys on the keyboard.
- If this situation appeared on the screen, what would you do?



Detect and Respond to Lights Alerts & Warnings (cont.)

- When you see a **yellow warning light** in the **2nd** Indicator Light Display (Fire Support), you should respond by pushing …

"2"

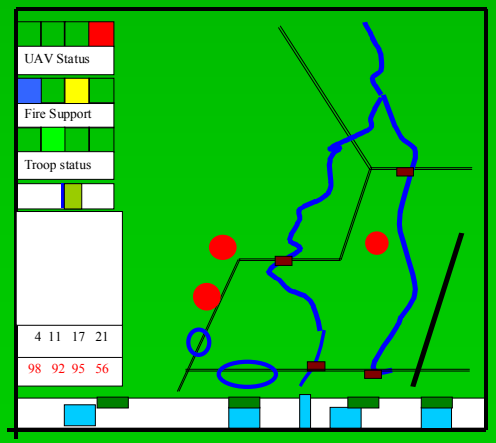


= push 2



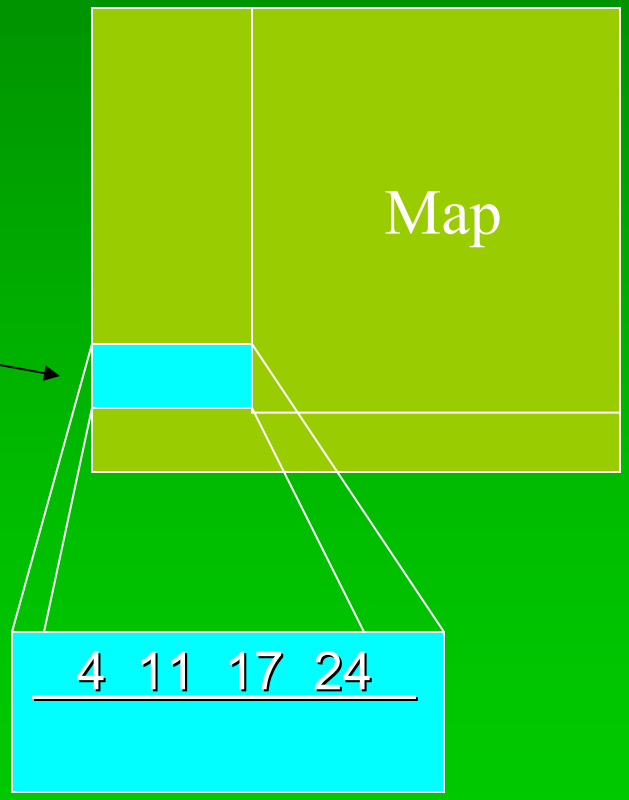
Go to the 2nd computer & run the VIKING program

(Press Ctrl F9 to restart Viking)

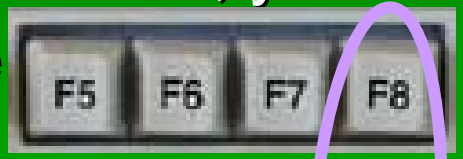


Links

- Detect link alert and evaluate numerical data
- The Link info is found in the lower left corner of the display
- Normally there are no numbers in the space below each link number (4 11 17 24)
  - But if there are #'s below, the system has detected a link failure.
  - You need to re-activate the link.



Links (cont)

- All #'s (i.e. 98 91 94 43) below the links should be ≥ 90
  - If they are NOT ≥ 90, you need to re-activate the links by:
    - Pushing the 

keys accordingly

4	11	17	21
98	91	94	43

    - In this case we would push the 

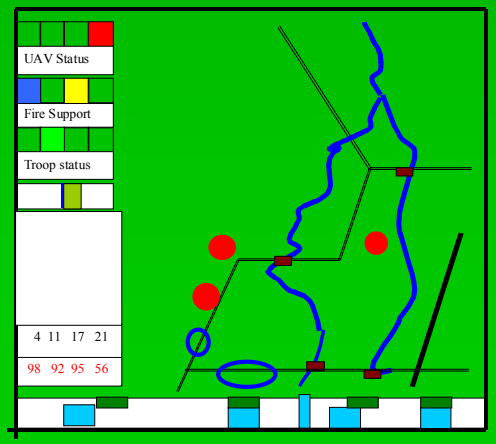
key because we see that 43 ≤ 90
  - If they ARE ≥ 90, you need to indicate a false alarm by:
    - pushing the 

key
    - The link will remain on the screen until the correct key is pressed



Go to the 2nd computer

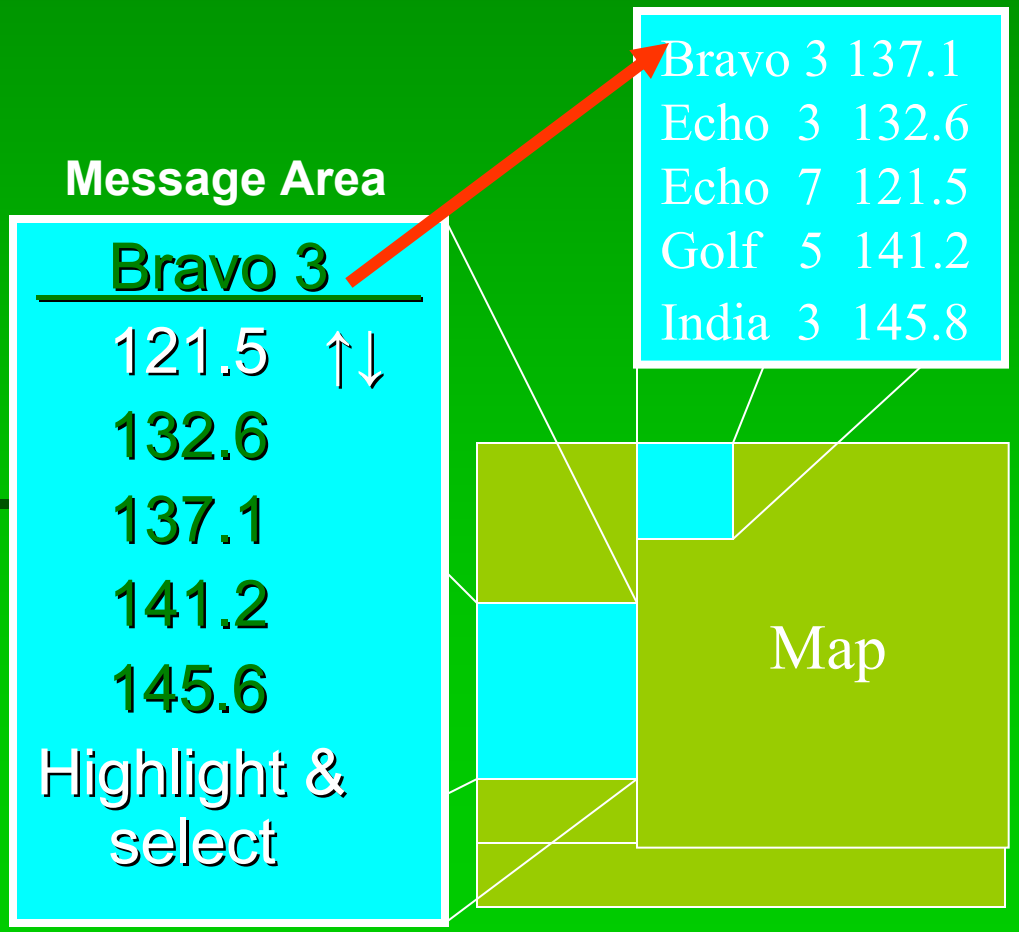
(Press Ctrl F9 to restart Viking)



Radio Frequency changes

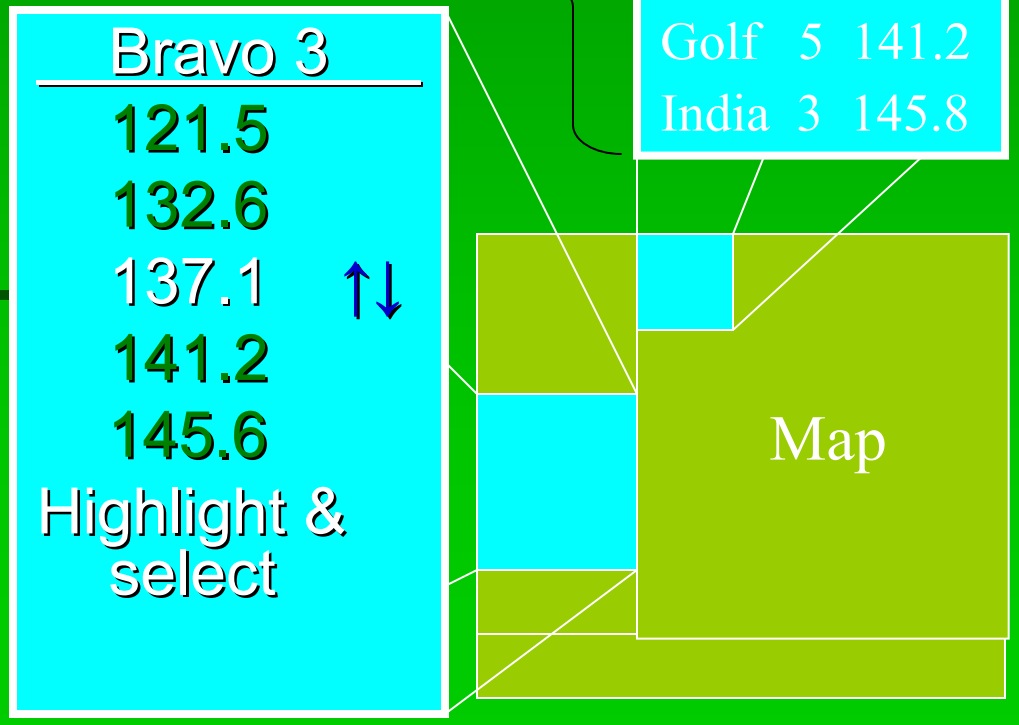
- When a frequency change is ordered, the code for the new freq and a listing of the available freqs appear in the message area.
  - The top freq (121.5) will be highlighted
- Response:
  - Use the ↑ and ↓ to move the highlight to the freq corresponding to the new code.
  - Refer to freq table

Frequency Table: Correct frequency codes are consistently present in the upper left corner of the Tactical Display



Radio Frequency changes (cont)

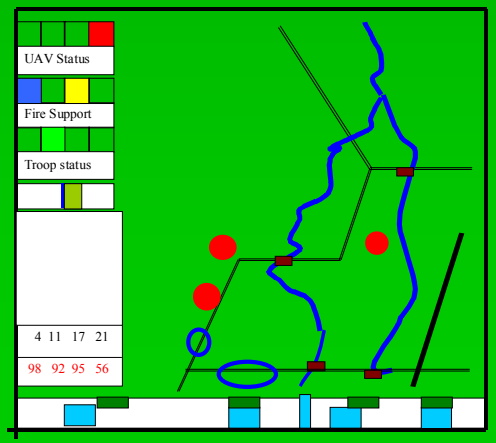
- Possible radio freqs are:
  - Bravo 3
  - Echo 2
  - Echo 7
  - Golf 5
  - India 3
- For this situation, Bravo 3 is the desired radio freq; and the corresponding frequency is 137.1 (see Freq Table)
- Response:
  - Use the ↓ to move the highlight to 137.1
  - Once at the desired freq, push the space bar to finalize your selection





Go to the 2nd computer

(Press Ctrl F9 to restart Viking)



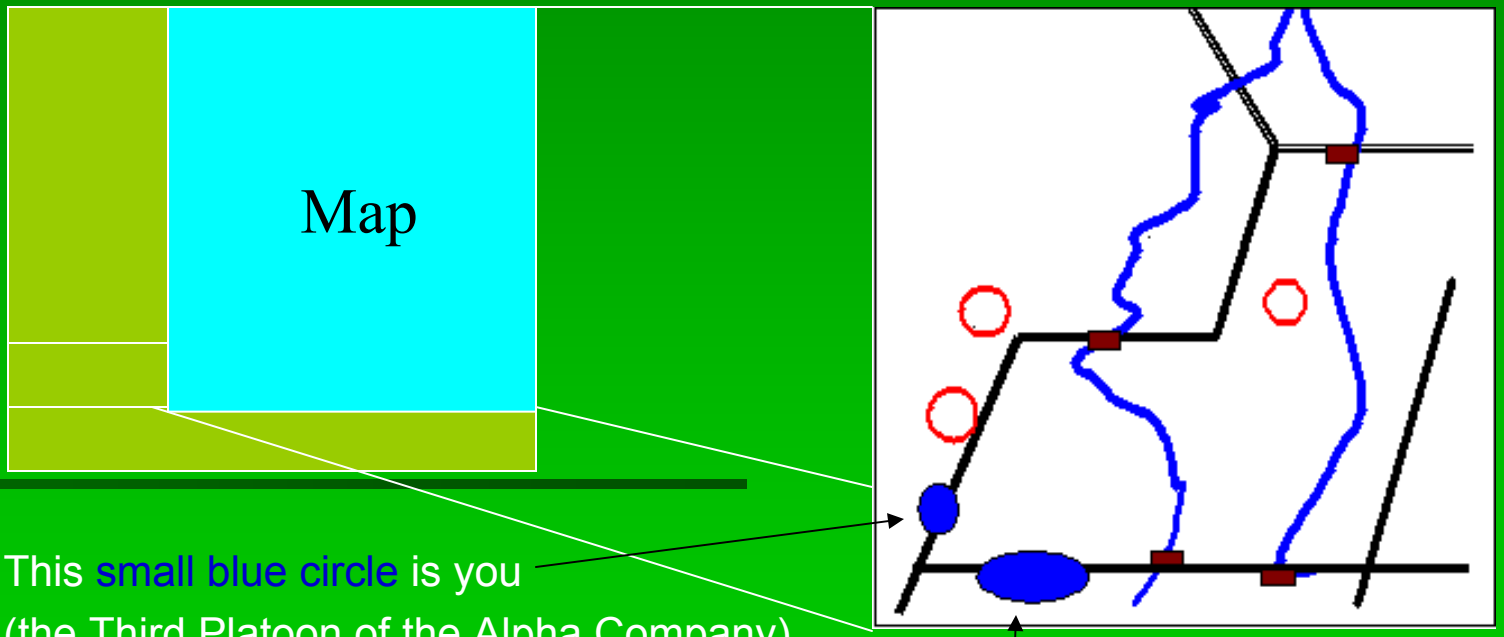
Platoon Spacing

- There are 2 components to the platoon movement / spacing task

1. Maintain proper spacing between platoon elements.
2. Maintain sufficient movement speed to reach rally point on time.

Platoon Spacing (cont.)

- Refer back to the screen layout

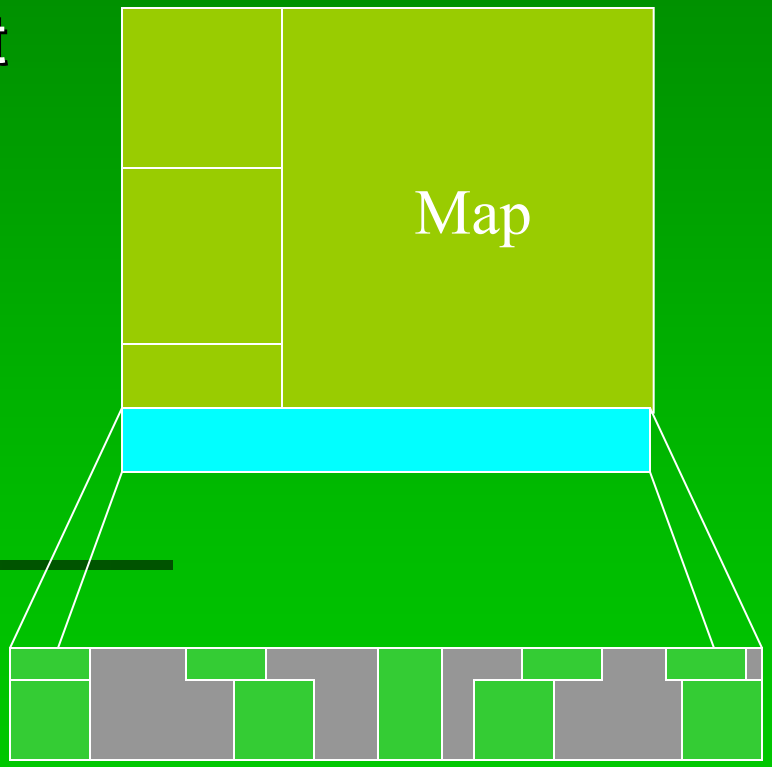


- This small blue circle is you
  (the Third Platoon of the Alpha Company)
  - The circle will visually move from the lower left corner to the upper right corner of the map as you control the Platoon Leader
- The large blue circle is the Alpha Company
  - You do NOT control the Alpha Company
  - The Alpha Company will NOT move

Platoon Spacing

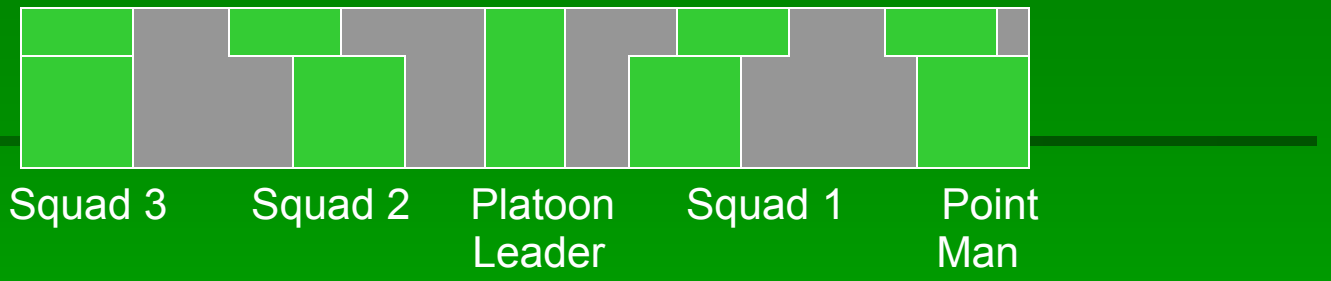
1. Maintain proper spacing between platoon elements

- The “platoon element position display” is found along the bottom of the display/screen
- This display indicates the location of the platoon units (4 squares) in relation to the platoon leader (vertical green bar)



Platoon Spacing (cont.)

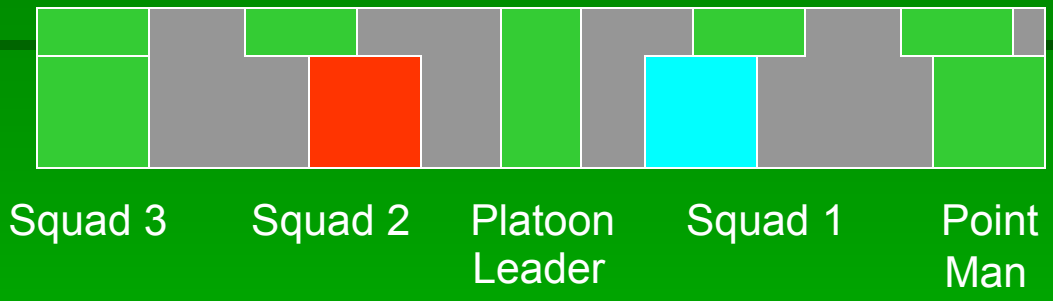
1. Maintain proper spacing between platoon elements



- You control all units of the platoon
- Movement
  - Squad 3, 2, 1 & Point Man WILL visually move on the display
  - the Platoon Leader (PL) will NOT visually move on the display
    - The Platoon leader is the paceman, the PL will always stay in the middle of the display
    - The PL may be stopped or started to maintain proper pace
  - The small rectangles along the top of the display will NOT move
    - they indicate the ideal location of the platoon units
- Color Change
  - Squad 3, 2, 1, & Point Man may change color
    - green, red, or light blue
  - The Platoon Leader may change color
    - green, or light blue
  - The small rectangles will NOT change color
    - They will remain green

Platoon Spacing (cont.)

1. Maintain proper spacing between platoon elements



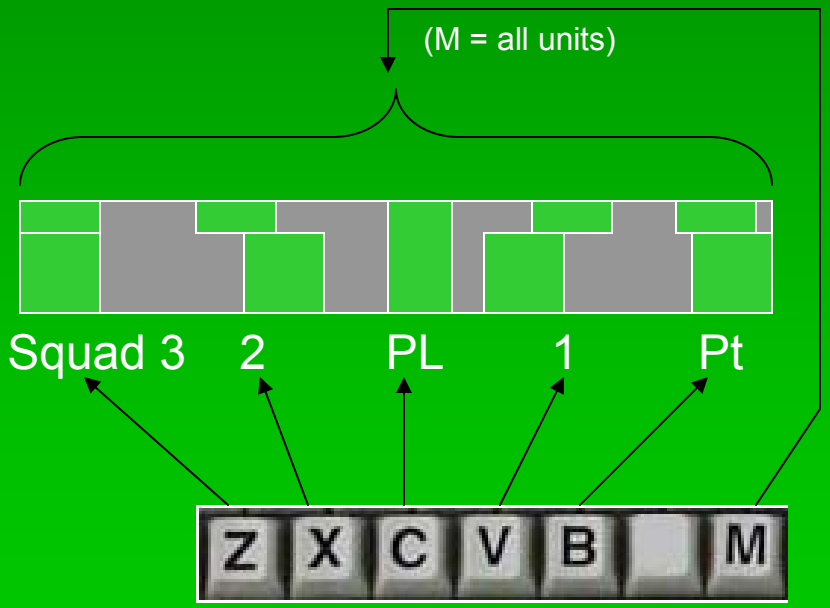
- The color of each square (Squad 3, 2, 1, and Point man) is important
- If the square is . . .
  - GREEN = the unit is MOVING
  - RED = the unit is UNABLE TO MOVE
  - LIGHT BLUE = the unit is STOPPED
  - A unit will always turn to blue after it has been red

Platoon Spacing (cont.)

1. Maintain proper spacing between platoon elements

- Each unit is programmed to move at different rates, thus stopping and starting units is necessary to maintain proper spacing

- To Start & Stop . . .
  - Squad 3 = push Z
  - Squad 2 = push X
  - Platoon Leader = push C
  - Squad 1 = push V
  - Point Man = push B
  - Entire platoon = push M



  - The "Entire Platoon M command does NOT work when any unit is disabled (RED). Units must be directed individually when any unit is disabled.

Platoon Spacing (cont.)

1. Maintain proper spacing between platoon elements

- If your display looks like this …



  - . . . You should stop Squad 2 (push X)
    - The GREEN square will turn LIGHT BLUE



  - You should stop Squad 2 because . . .
    - You don't want Squad 2 to march too far ahead of the other units
    - Squad 1 is unable to move & will continue to fall behind schedule
    - When Squad 1 is ready to move again, the square will change from RED to LIGHT BLUE.



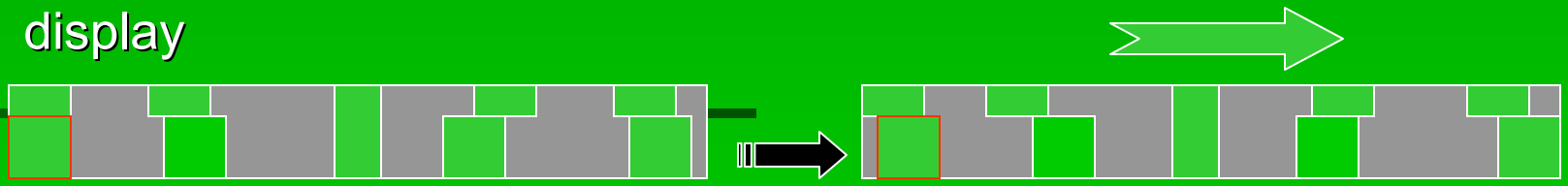
      - At this time you should push V to make Squad 1 move again
      - the square will change from LIGHT BLUE to GREEN



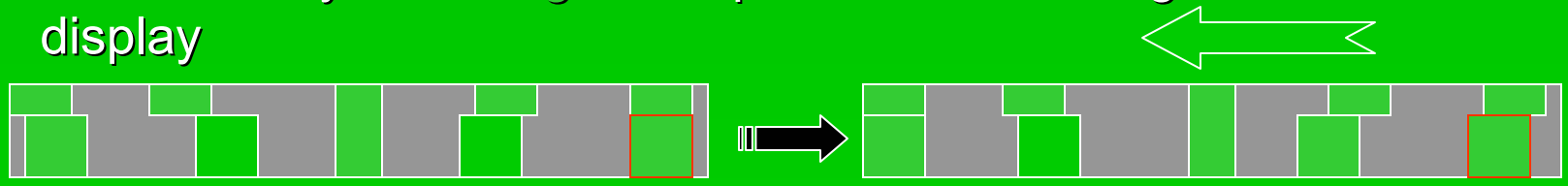
Platoon Spacing (cont.)

1. Maintain proper spacing between platoon elements

- FYI: As time continues, (green) units will continue to move (march)
  - Some units may keep pace with the Platoon Leader, thus we will NOT visually see the green square move on the display
  - Some units may march faster than the Platoon Leader, thus we WILL visually see the green square move from left to right on the display



  - Some units may march slower than the Platoon Leader, thus we WILL visually see the green square move from right to left on the display



Platoon Spacing (cont.)

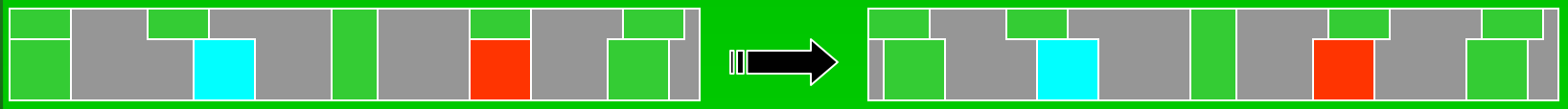
1. Maintain proper spacing between platoon elements

- FYI: If a unit is either red or blue . . .
  - The unit is not moving
    - not marching & not moving visually on the screen
  - As time goes on . . .
    - the stopped unit will move from right to left on the screen;



    - as green units continue to move left to right (at different speeds) on the screen





    - To help a unit “catch up”, it may be necessary to stop the platoon leader and thus the platoons rate of movement

Platoon Spacing

2. Maintain sufficient movement speed to reach rally point on time

- the “Time/position Indicator” begins with a red line in the middle of the indicator (On schedule)



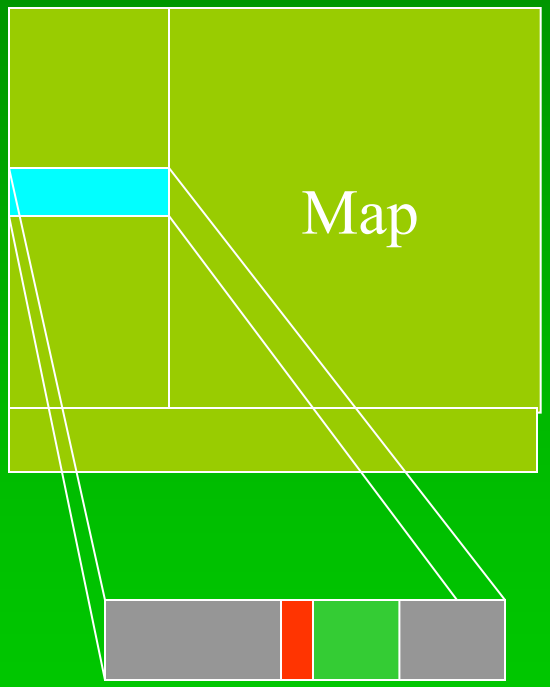
- If the platoon is ahead of schedule, a green bar moves to the right (from the red line)



  - Keeping the platoon moving with minimal stops will help stay on & move ahead of schedule
- If the platoon is behind schedule, a red bar moves to the left (of the red line)



  - Stopping the column & adjusting spacing takes time



Time/position Indicator

Platoon Spacing (cont)

2. Maintain sufficient movement speed to reach rally point on time

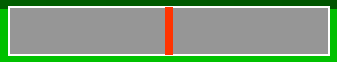
- The longer the bar is (in either color/direction), the further the platoon is ahead of or behind schedule (whichever applies)
  - Extremely behind schedule



  - Slightly behind schedule



  - On time



  - Slightly ahead of schedule



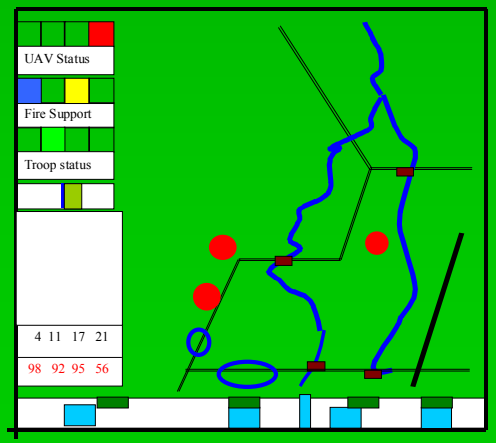
  - Extremely ahead of schedule





Go to the 2nd computer

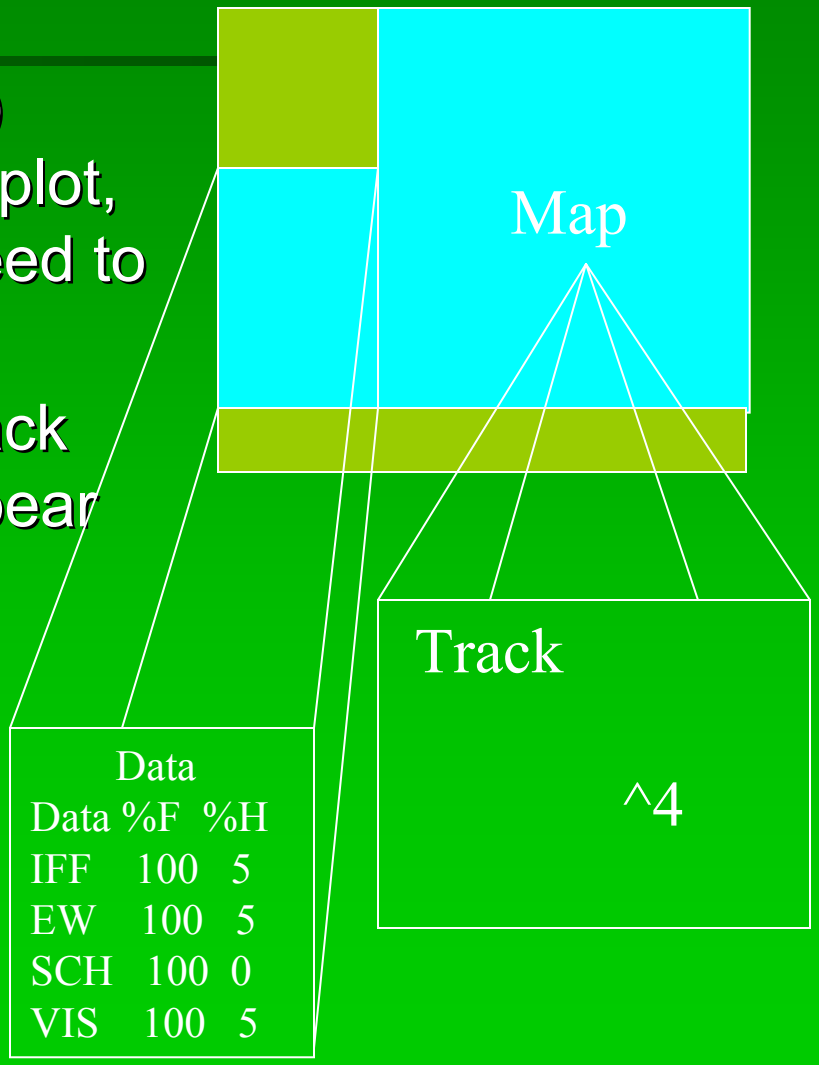
(Press Ctrl F9 to restart Viking)



Aircraft Identification

- When a track (i.e. ^4, ^9, etc.) appears on your map/tactical plot, there is an aircraft that you need to identify
- Almost immediately after a track becomes visible, data will appear in your message window
- From this data, you need to determine if the unidentified aircraft is:
  - Friend

    - or -
  - Foe/Hostile



Aircraft Identification (cont)

- A closer look at the data . . .

Data	%F	%H
IFF	100	5
EW	100	5
FP	100	0
VIS	100	5

Data	%F	%H
IFF	75	25
EW	100	25
FP	75	0
VIS	50	25

Data	%F	%H
IFF	50	50
EW	100	100
FP	50	0
VIS	50	50

Data	%F	%H
IFF	25	75
EW	25	100
FP	25	0
VIS	25	25

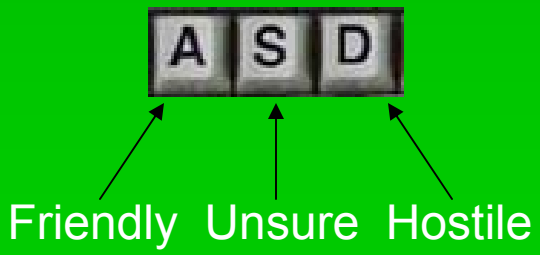
Data	%F	%H
IFF	5	100
EW	5	100
FP	0	100
VIS	5	100

Friendly → Unsure → Hostile

- Data:
  - IFF = Interrogation Friend or Foe
  - EW = Electronic Warfare
  - FP = Flight Path
  - VIS = Visual & Auditory Data
- %F = Percent Friend
- %H = Percent Hostile/Foe
- Data entries in the left most column strongly support a Friendly aircraft classification
- As the examples continue from left to right, the data become more supportive of a Hostile aircraft classification

Aircraft Identification (cont)

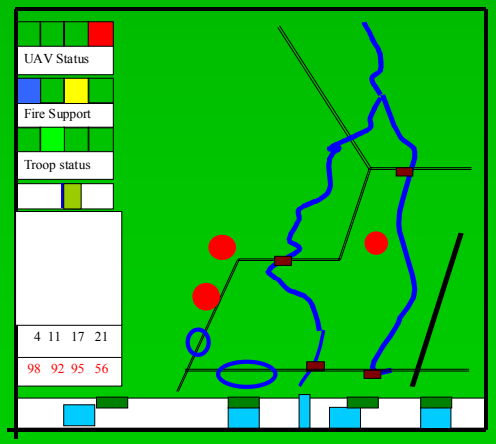
- From the data given, you need to respond in 1 of 3 ways:
  1. If you think the track is **Friendly** = push A
  2. If you think the track is **Hostile** = push D
  3. If you cannot classify (**unsure** if it is friendly or hostile) = push S



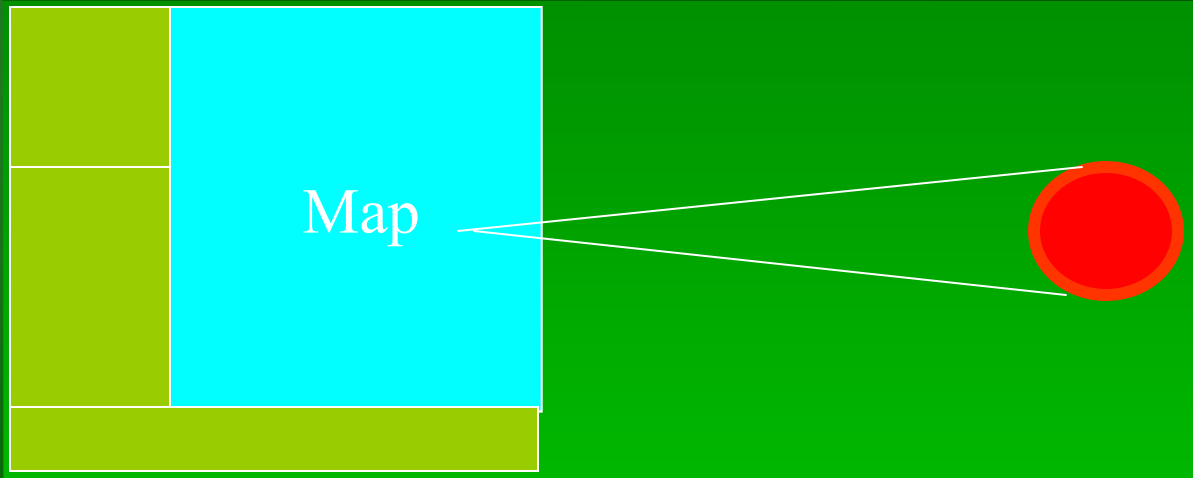


Go to the 2nd computer

(Press Ctrl F9 to restart Viking)

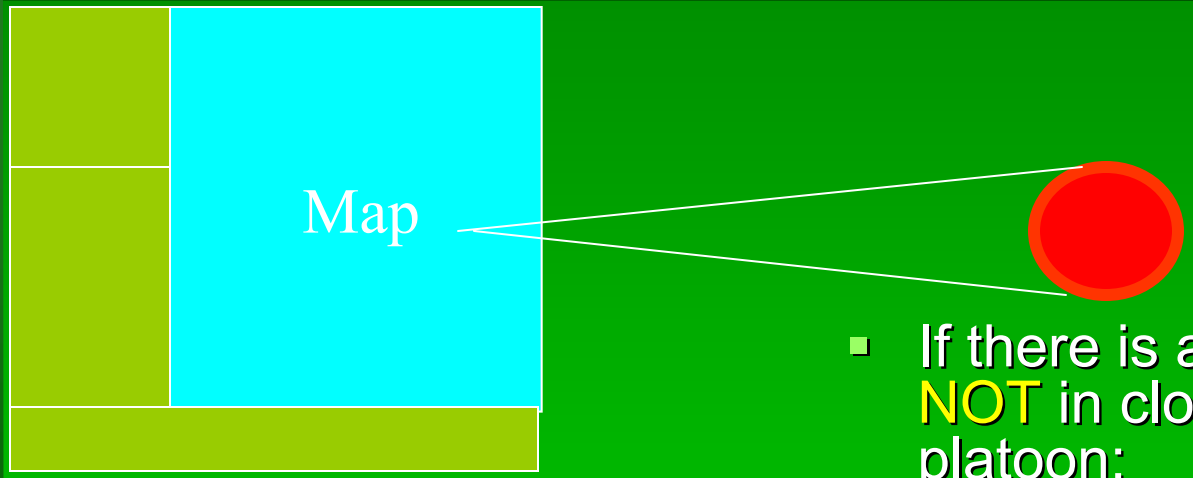


Detect & Destroy Enemy



- If there is an Enemy Unit in the map area, the Enemy Unit will appear as a red circle
- It is possible to have more than one Enemy Unit on the map

Detect & Destroy Enemy



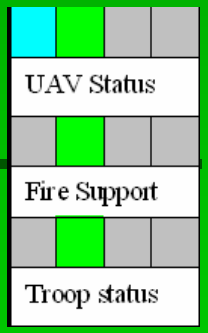
- If there is an enemy unit that is in close proximity to your platoon:
  - the enemy unit will appear automatically on the map as the platoon makes contact

- If there is an enemy unit that is NOT in close proximity to your platoon:
  - The enemy unit may automatically appear by routine UAV fly-overs

    - or -
  - The enemy unit will appear after you request a UAV fly-over
  - You may request UAV fly-overs as often as you like

Detect & Destroy Enemy (cont)

- Enemy forces =
- Responses & Info:
  - UAV Launch = push 0 (zero)
    - A blue light will appear in the “UAV Status” Light Display
      - Indicating that the UAV is gathering data



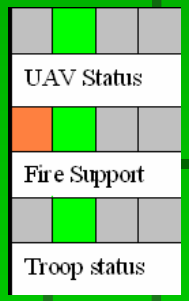
    - If Enemy Units are detected, a red circle will appear on the map
    - If Enemy Units are NOT detected, nothing will appear on the map
    - The blue UAV Status Light will automatically disappear in approximately 5 seconds regardless of the result of the UAV launch

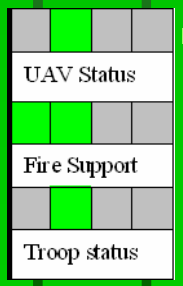
Detect & Destroy Enemy (cont)

- Responses & Info (cont):
  - Enemy force targeted = push /
    - Once you have pushed /, the RED circle will turn **WHITE** (highlighted)



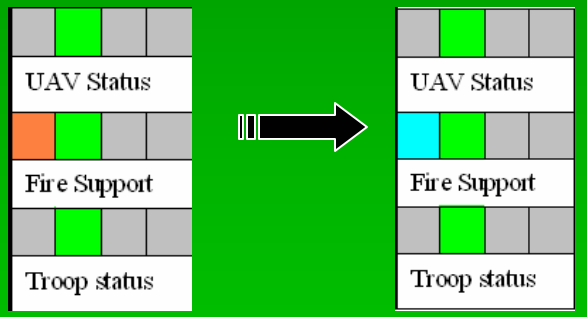
    - If there are more than 1 enemy force shown on the screen, pressing / will cycle thru each enemy force (turning each red circle to **white** in turn)
    - An orange light in the left most square in the "Fire Support" Light Display will appear when an Enemy Force has been targeted (**white** circle)
    - If you push / again, the highlight will be eliminated, no Enemy Forces will be targeted
      - All enemy forces will be red circles
      - the red light will turn green (safe) in the left most "Fire Support" Light



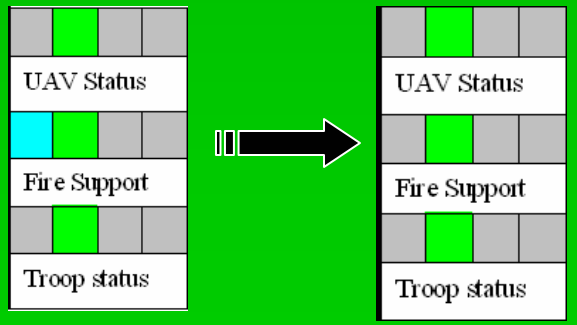


Detect & Destroy Enemy (cont)

- Responses & Info (cont):
  - When an Enemy Unit has been targeted (**white** circle), you can destroy it
  - Engagement = push P
    - The left "Fire Support" light changes from orange to blue (mission accepted)

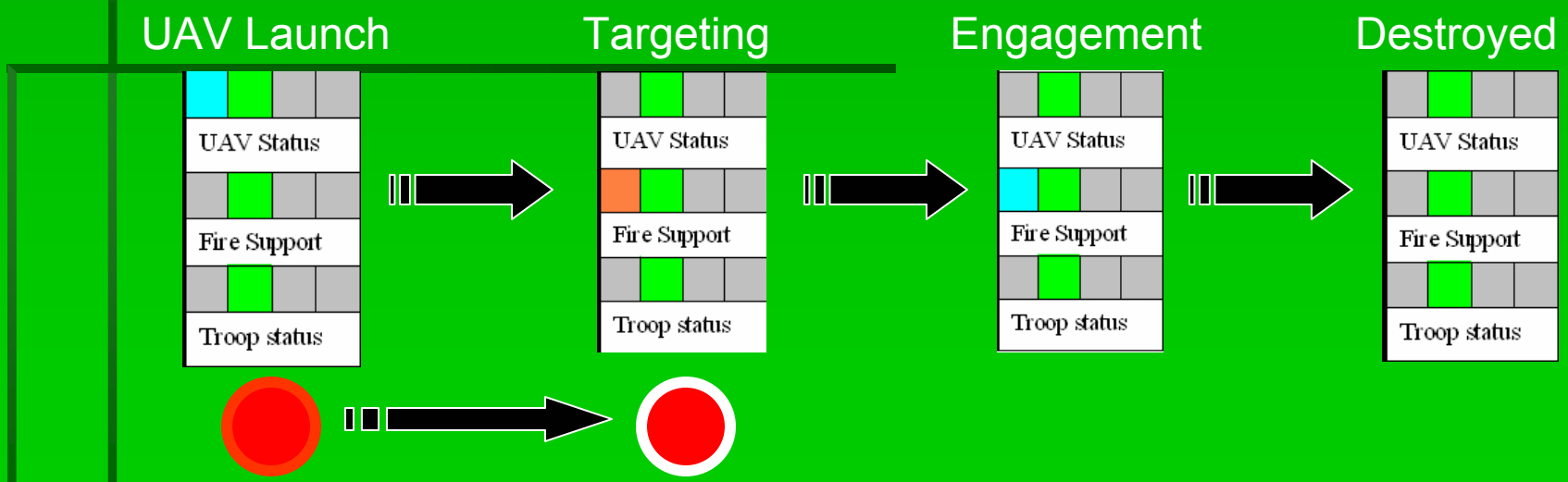


    - After approx 5 seconds . . .
      - the highlighted target disappears
      - the blue fire support light disappears



Detect & Destroy Enemy (cont)

- Recap
  - Responses:
    - UAV Launch = push 0
    - Targeting = push /
    - Engagement = push P

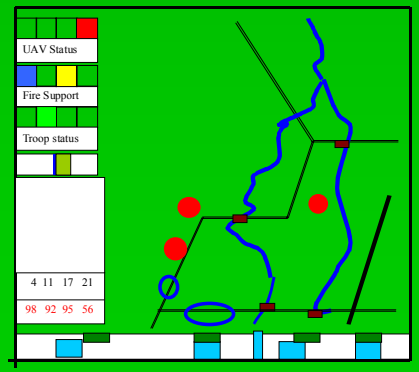


End of PowerPoint Presentation



Go to the 2nd computer

(Press Ctrl F9 to restart Viking)



Stress and Cognitive Performance Field Assessment

Quarterly Report
31 December 2002

Wayne C. Harris
Human Factors Research Laboratory
University of Minnesota

Overview

Activities during the fourth quarter of 2002 included: 1) analysis of field data collected at Ft. Chaffee, 2) attendance at the Annual meetings of the Human Factors and Ergonomics Society meetings and a MURI meeting in Orlando, 3) analysis of the mental workload and cognitive performance data, 4) a visit and presentation at the U.S. Army Research Institute for Environmental Medicine at Natick MA, 5) preparation for data collection at the Marine School of Infantry (West) at Camp Pendleton, and 5) continued development of military field training data collection opportunities.

Ft. Chaffee field data analysis

Analysis of field data was continued and the manuscript for submission of the results for presentation at the 2003 Human Factors meetings was begun.

Data analysis for the mental workload study

To determine whether mental effort affects the measures of cognitive performance, an experiment was conducted which examined the effect of three levels of workload on the Viking task. Data included salivary amylase, ARES cognitive battery performance, and Viking multi-task performance. This study will also provide background for the use of the Viking task in field studies. A paper based upon this study will be submitted to Human Factors for presentation at the 2003 meetings.

Preparation for data collection at the Marine School of Infantry (West) at Camp Pendleton

The forms required to receive human subject approval are being completed and the specific research protocol is being finalized as a part of this process.

Development of future field data collection sites

Two sites for future studies have been identified. One is the Special Forces training center at Ft. Bragg, NC and the second is the Survival, Evasion, Resistance and Escape School at Brunswick ME. Points of contact have been identified and initial contacts have been made.

Stress and Cognitive Performance Field Assessment

Quarterly Report
30 September 2002

Wayne C. Harris
Human Factors Research Laboratory
University of Minnesota

Overview

Activities during the third quarter of 2002 included: 1) analysis of field data collected at Ft. Chaffee, 2) presentation of Ft. Chaffee data at the Annual Meeting of the American Psychological Association in Chicago, 3) collection of cognitive data in the laboratory to assess the effect of mental workload on the cognitive battery being used in the field and to examine the effect of mental workload on salivary amylase, 4) a presentation to the Marine School of Infantry (West) at Camp Pendleton to obtain permission to collect data during live-fire exercises, and 5) initial contacts with additional military field training sites regarding data collection opportunities.

Study: Ft. Chaffee field exercises

An initial analysis of field data was completed and the results were presented at the Annual Meeting of the American Psychological Association in Chicago. A more detailed analysis is currently being performed.

Study: "The effect of mental workload on cognitive performance and salivary hormone"

Physical effort, mental effort and danger are among the factors that may affect performance and salivary hormones in field settings. Dr. Leon's laboratory is examining the effect of physical effort. To determine whether mental effort affects the measures of cognitive performance we are using in the field, an experiment was conducted which examined the effect of three levels of workload. Mental workload was produced by the Viking program, which had been developed as a part of the MURI program. Data has been collected and analysis has begun. The experiment also provided experience with the Viking and with salivary hormone collection, skills which may be needed during future field studies. A PowerPoint Viking training program was developed a part of the study.

Site Development: Marine School of Infantry (West) at Camp Pendleton

A briefing was provided to the Commanding Officer of the Marine School of Infantry (SOI). He indicated that he will support data collection during live-fire exercises at the SOI. Observation of a live-fire exercise and a discussion of the details of data

collection with the Executive Office occurred the following day. Specific dates and procedures are currently being discussed.

Contacts with additional field sites

Contacts have been made with Army Research Laboratory personnel at Ft. Benning to identify potential data collection opportunities. A PowerPoint briefing was developed and forwarded to ARL personnel at the site.

Stress and Cognitive Performance Field Assessment

Quarterly Report
29 March 2002

Wayne C. Harris
Human Factors Research Laboratory
University of Minnesota

Overview

Activities during the first quarter of 2002 included completion of the infantry multi-task experimental environment (Viking), testing and installation of the cognitive battery on field-hardened computers in preparation for field-testing, and administrative activities including collaborative interaction with other MURI sites and military units.

Development of an infantry multi-task experimental environment

1. Completion of an infantry multi-task environment, the Viking
    a. The first generation of infantry multi-task environment was completed early in the quarter.
    b. First generation software was evaluated and data collection was added for new tasks.
        1) The Generation One program was tested in our laboratory.
        2) The Generation One program was shared with other MURI laboratories, which provided feedback that was incorporated into the program's evaluation.
        3) Baseline data was collected during repeated administration to evaluate skill acquisition.
        4) Data collection capability was added for the tasks that had been added to the final version of the Generation One program.
        5) A generation Two program was created based upon the results of the Generation One program evaluation.
    c. A training program was developed that includes a PowerPoint instructional program and a practice session.
    d. A data analysis program was developed that extracts and summarizes Viking data files. The program was then shared with other MURI laboratories and changes were introduced into the program based upon the feedback from those laboratories.
    e. Presented a paper at the Mid-year Meetings of APA Divisions 19 and 21 in Ft. Belvoir, VA. The presentation provided an overview of the Viking program and indicated potential uses for the program.

f. Documentation of the Viking program and the Viking data summary program was updated to reflect program changes and additions.

Testing and installation of the cognitive battery in preparation for field-testing.

1. MURI cognitive battery has been installed on Toughbook computers in preparation for field-testing.
2. An evaluation is underway with personnel at Camp Pendleton regarding the use of a hand held testing device currently being used to evaluate U. S. Army soldiers in Bosnia. Use of this device would facilitate comparison of data with ongoing Army projects.
3. The laboratory supported the cognitive assessment efforts of the other MURI laboratory at the University of Minnesota. Results of that effort include a presentation at the Mid-year meeting of APA Divisions 19 and 21 at Ft. Belvoir, VA.

Administrative tasks

1. *Field testing site access*
   Support from the MN National Guard was obtained and efforts to identify specific units and training activities are now underway. Activities have moved to identifying specific units and contacting key personnel in those units. Current discussions are focusing on Ft. Chaffee during June and Ft. Irwin in September.

2. *Staffing*
   A summer intern has been hired. The intern is a graduate student in the Industrial/Organizational Psychology and Human Factors program at Minnesota State University, Mankato. Her background is well suited for the project as she is a Lieutenant in the Army National Guard where she serves as a helicopter pilot.

3. *Support activities*
   The laboratory supported the data analysis of other MURI sites using the ANAM.

4. *Collaboration*
   Interaction with other MURI sites continued during this period remained a collaborative active. Activities associated with MURI development included a visit to the Cognitive Science Laboratory at Catholic University of America.

5. *Resource development*
   Defense Technical Information Center registration was completed. Access to previous DoD results will increase the availability of data relevant to the effect of combat stress.

6. *Information dissemination*
   Two papers were presented at the Mid-year Meetings of APA Divisions 19 and 21 in Ft. Belvoir, VA. The papers summarized the development of the Viking and examined the data obtained in Dr. Leon's laboratory at the University of Minnesota that assessed the effect of exercise on cognitive performance.

MURI Interim Progress Report

Report Period: 1st Quarter 2002 - March 29, 2002

To: Dr. Peter Hancock, Principal Investigator, University of Central Florida

From: Drs. Arthur Leon, Principal Investigator & Ava Walker, Project Coordinator
University of Minnesota
Laboratory of Physiological Hygiene and Exercise Science

Re: MURI Subcontract Agreement w/University of Minnesota
under Army #DAAO –19-01-1-0621
"Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield"

1. Overview. The activity for the MN Center during the first quarter of Phase 2 (01Jan02 - 31March02) included protocol revision and Manual of Procedure adjustments, recruiting strategies were revised around a lower participant compensation plan and, institutional review board approval was revised and resubmitted. Budget issues involving a computer purchased delayed data collection for much of this 1st Quarter. Laboratory analysis was completed for salivary amylase and cortisol, and data entry was completed for 160 samples. The change in the semester brought new undergraduate and graduate students to the project. Database adjustments were made and a preliminary analysis of the data provided enough information for the production of three abstracts presented to the Annual Midyear Symposium of the American Psychological Association, Division 19 & Potomac Chapter, Human Factors & Ergonomics Society March 7-8th, 2002 at Fort Belvoir, Virginia. Presentations were made by Dr. Ava Walker and Dr. Wayne Harris. This invitation also provided an opportunity to have most of the MURI players together for an update meeting.

The following tasks were accomplished during 1st Quarter Phase 2:

2. Organization and Logistics tasks. Task that were completed include:
   a. administrative tasks related to IRB revision/resubmission
   b. computer ordered but not received in this quarter
   c. student staff orientation and training
   d. procedures modified with U of MN/Fairview Medical Center Biochemistry Laboratory for salivary tests – results printouts were not being delivered in a timely manner

3. Protocol and Assessment Technique Revised.
   a. Speed of treadmill walking increased from 3.4mph to 3.7mph to facilitate a greater increase in heart rate with some relief in percent grade on the treadmill

4. Abstracts presented at the Annual Midyear Symposium of the American Psychological Association, Division 19 & Potomac Chapter, Human Factors & Ergonomics Society March 7-8th, 2002 at Fort Belvoir, Virginia

A. J. Walker, W. C. Harris, A. S. Leon, & Peter Hancock, Ph.D. Changes in salivary amylase

and self-reported fatigue across three increasingly intense bouts of exercise.

A. J. Walker, H. L. Bolgren, K. G. Nichols, A. S. Leon & Peter Hancock, Ph.D. Salivary amylase as a marker of increasing physiological stress during a 5-mile treadmill road march.

W. C. Harris, A. J. Walker, A. S. Leon, & Peter Hancock, Ph.D. The effect of physical workload on cognitive performance.

PowerPoint Slide Outline Format

Salivary amylase as a marker of increasing physiological stress during exercise

A. J. Walker, H. L. Bolgren, K. G. Nichols, & A. S. Leon University of Minnesota

And Peter Hancock, Ph.D. University of Central Florida

1) Physical Activity –
- acute, intense bouts
- chronic habitual activity
- psychological benefits
- increases in academic performance,
- assertiveness
- confidence
- emotional stability (Thomas, Landers, Salazar, Etnier, 1994

2) Objectives
- To examine the effects of two different exercise challenges on physiological fatigue, ratings of perceived exertion, and cognitive function

3) Method
- Participants
- Physiological assessment
- Procedures

4) Subjects
- 3 females and 7 males
- 19 to 38 years of age (Mean = 24.4)
- Military personnel
  University of Minnesota R.O.T.C.
  National Guard
  Active duty

5) Profile of Mood State

$VO_{2\ Max}$ Test

Intensity Testing 30 – 60 – 90% $VO_{2\ Max}$

Intensity Testing 30 – 60 – 90% $VO_{2\ Max}$

6) Cognitive Assessment
- Automated Neuropsychological Assessment Metric (ANAM)
  6 sub-tests
- Test selection criteria
  Minimize psychological demand overlap task
  Sample critical information processing components

7) Sub-test Selection
- ☐ Information processing stages (Walter Reed PAB)
- ■ Perceptual input, detection & identification
- ■ Central processing
- ▫ Short- and long-term memory, associative memory
- ▫ Information integration/manipulation
- ▪ Linguistic/symbolic
- ▪ Spatial

8) Sub-tests
- ☐ Simple reaction time
- ☐ Coding
- ☐ Match to sample
- ☐ Memory (2 and 6 item)
- ☐ Spatial processing
- ☐ Logical reasoning

9) PROCEDURES
Training
Treadmill

10) Intensity
- ▪ Low intensity (30%)
- ▪ Moderate intensity (60%)
- ▪ High intensity (90%)

11) Five - mile march
Percent Changes in the Fatigue Score in POMS
and Salivary Amylase During Road March

12) Percent Changes in the Fatigue Score in POMS
and Salivary Amylase During Road March

13) **ACKNOWLEDGEMENT**

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

Respectfully Submitted,

Arthur S. Leon, MD Ava J. Walker, Ph.D

Powerpoint Slides enclosed w/email

The Effect of Trust on Team Problem Solving

Nathan Fosse
Lauren Murphy
Jessica Perry
Liz Ratcliff
Kip Smith

Outline

- Introduction
  - Issues
  - Hypotheses
- Method
- Results
- Implications

Research question: Team performance

- General: Do different levels of trust affect performance on a cooperative problem solving task?

- Specific: Do trust, distrust, and betrayal affect team play on the Hobbits and Orcs river crossing problem?

Research question: Individual performance

- General: Do different levels of trust affect performance on a basic cognitive process?
- Specific: Do trust, distrust, and betrayal impede practice effects on mental rotation of simple 3-D objects?

Hypothesis: Team performance

- Trusting groups will perform best
- Distrusting groups will have an intermediate level of performance
- Betraying groups will perform worst

Hypothesis: Individual performance

- Levels of trust will not be detrimental to individual cognitive processes, e.g., mental rotation, spatial reasoning

- We expect practice effects in a repeated task
  - Performance should improve the second time around

Experimental design

1 Pretest of individual performance

2 Intervention - trust game

3 Team play - Hobbits and Orcs

4 Post-test of individual performance

Individual cognitive task

- Pretest - Intervention - Team play - Post-test

- Mental rotation
  - (Metzler & Shepard, 1972)
- A classic problem from cognitive psychology

Quasi-manipulation of trust

- Pretest - Intervention - Team play - Post-test

- Extensive form trust game
  - (McCabe, Rassenti, & Smith, 1996)
- Self-selection into one of three conditions
  - Distrusting
  - Trusting/trustworthy
  - Betraying

Cooperative problem solving task

- Pretest - Intervention - Team play - Post-test

- Hobbits and Orcs
  - (Greeno 1974)
- A classic problem from cognitive psychology

The expected dissociation

- Team:

Teamwork will respond to the quasi-manipulation of trust.

- Trust game
- Team play

- Mental rotation

- Individual:

- Mental rotation

Individual cognitive processes will NOT respond to the quasi-manipulation of trust. Practice will improve performance.

- Mental rotation

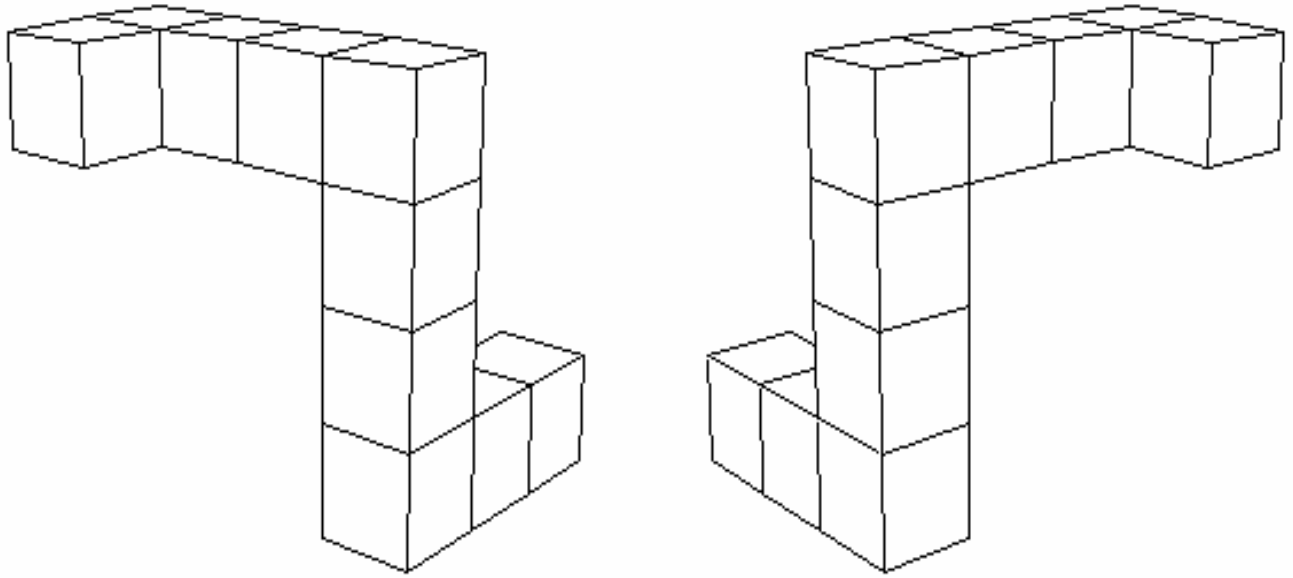
Method

- Subjects
- Mental rotation
- Trust game
- Hobbits and Orcs
- Analyses

Subjects

- 14 teams of 2
- Retain anonymity through play of the trust game
  - Sign consent forms
  - Mental rotation pretest
  - Trust game
- Come together
  - Trust game payout
  - Team play
- Individually
  - Mental rotation post-test

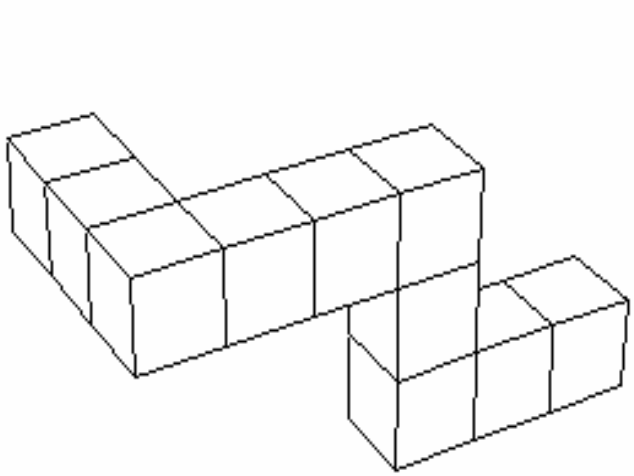
Mental rotation

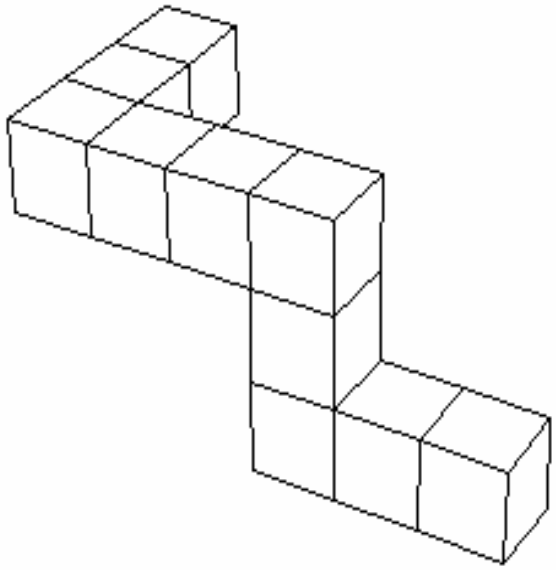


Are these figures the same or different?

Different

Mental rotation





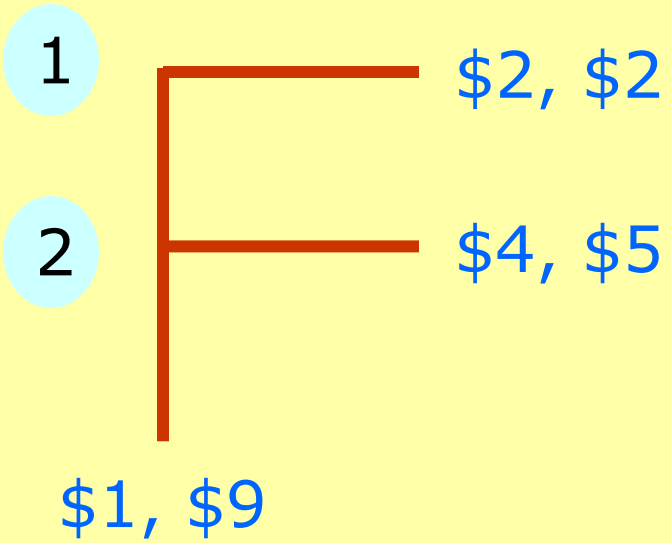
Are these figures the same or different?

Same

The trust game

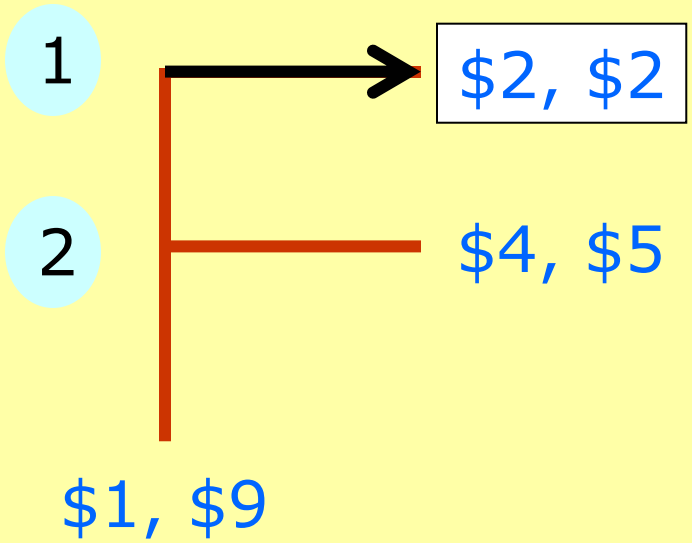
- The sequential, extensive-form two-person trust game

- The game has three possible outcomes
  - Distrusting
  - Trusting & trustworthy
  - Betraying



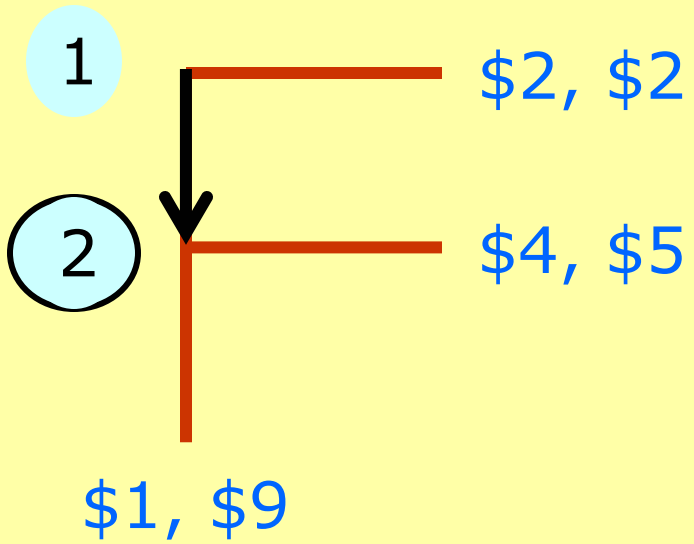
Outcome 1 - Distrust

- Sequential play
- Player 1 can move right
  - Player 1 gets $2
  - Player 2 gets $2
  - Player 2 doesn't get to play
  - Game over



Trust

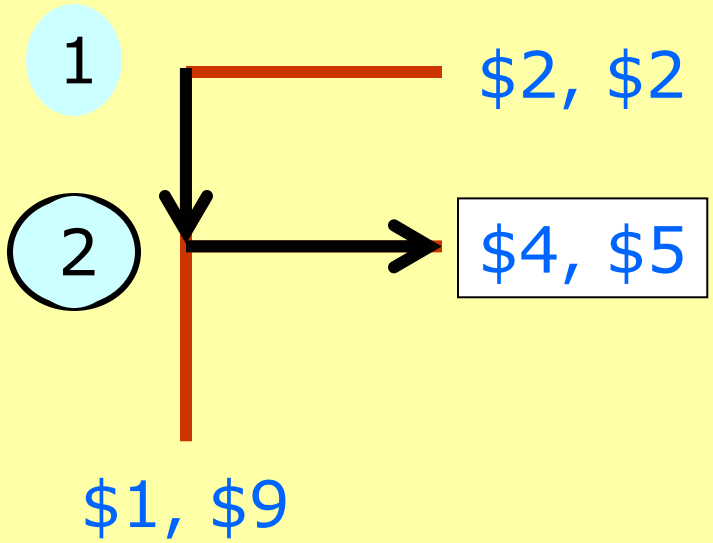
- Sequential play
- Player 1 can move down
  - Passing play to Player 2
  - An explicit act of trust



Outcome 2 - Trustworthy

- Passing play to Player 2 is an explicit act of trust

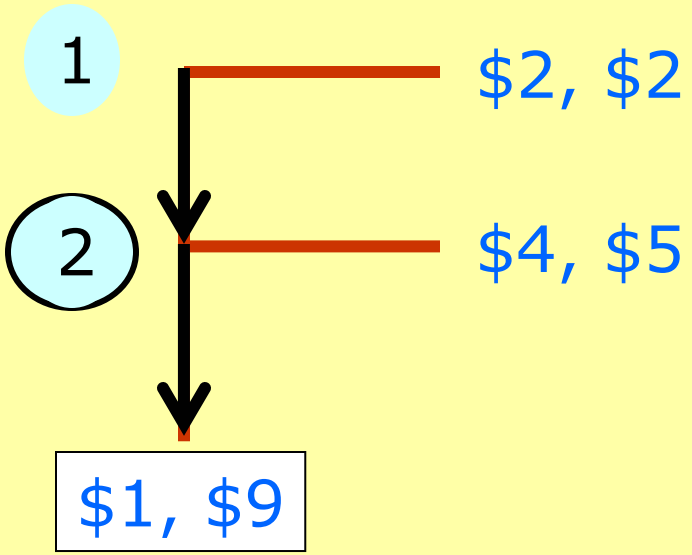
- If Player 2 then moves right, the trust is reciprocated; Player 2 is trustworthy
  - Player 1 gets $4
  - Player 2 gets $5
  - Game over



Outcome 3 - Betrayal

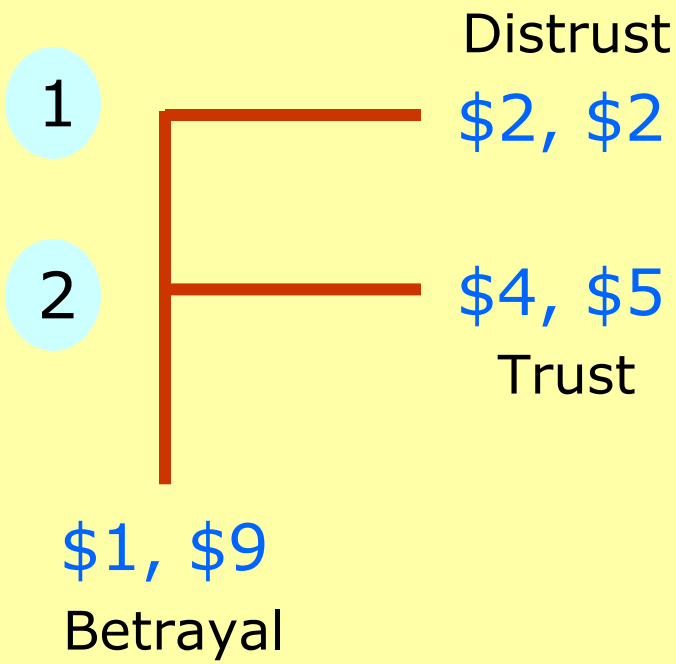
- Passing play to Player 2 is an explicit act of trust

- If Player 2 then moves down, it is a betrayal of Player 1's trust
  - Player 1 gets $1
  - Player 2 gets $9
  - Game over



Three outcomes: Three levels of the IV

- The moves made by subjects determine the team's level of trust = a quasi-manipulation
  - The IV is a subject variable
- In all previous studies of trust on performance, the experimenters have manipulated trust



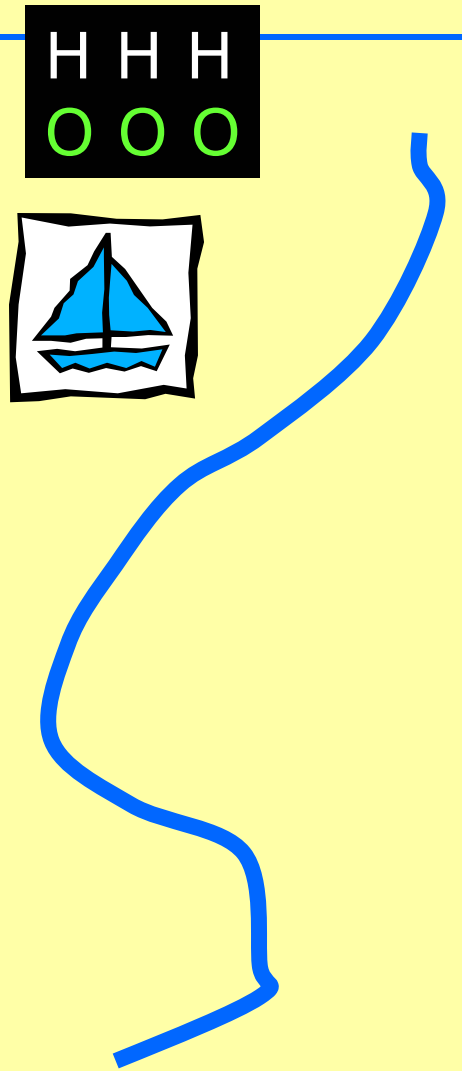
Restatement of the hypotheses

- Team performance (trusting) >
  Team performance (distrusting) >
  Team performance (betraying)

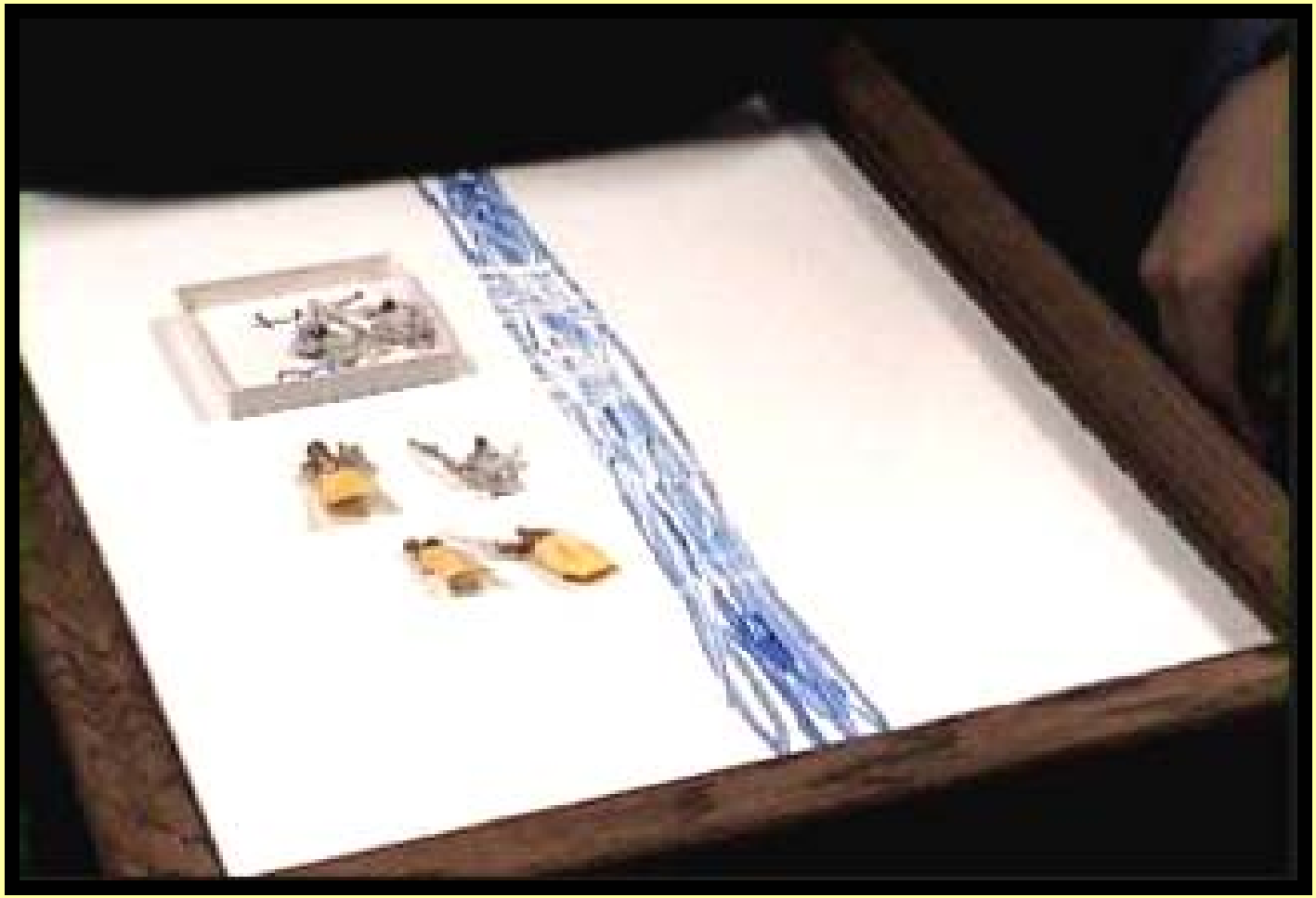
- For all conditions:
  Mental rotation in Post-test >
  Mental rotation in Pretest

Hobbits and Orcs

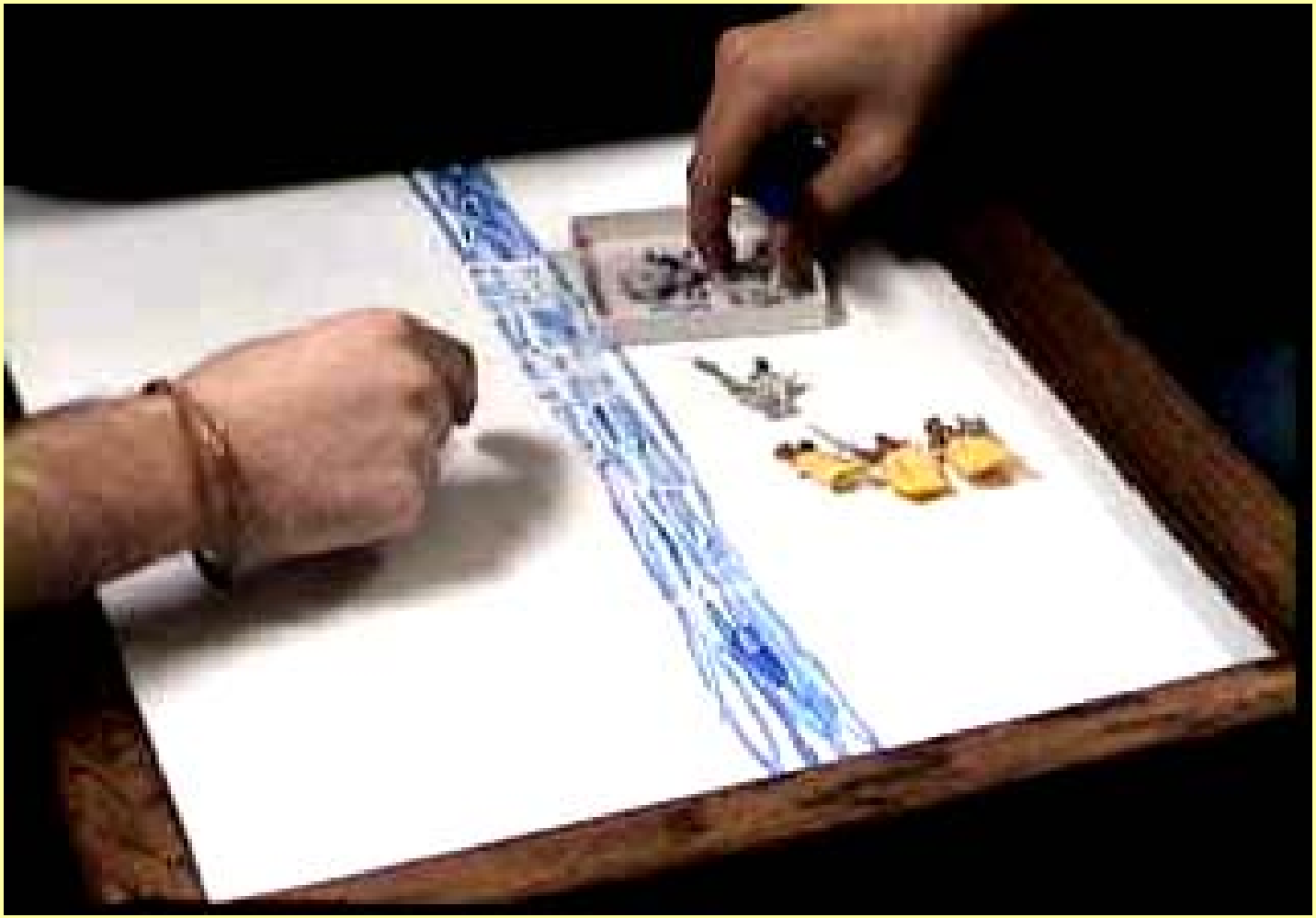
- Three hobbits and three orcs need to cross a river
- Rules:
  - Only two critters in the boat at any one time
  - Orcs cannot outnumber hobbits anywhere
  - Player 1 moves right
  - Player 2 moves left
  - 100 moves or 45 minutes



Hobbits and Orcs



Hobbits and Orcs



DVs and Analyses

- Between subjects (teams) ANOVA on performance on Hobbits and Orcs
- The team is the unit of analysis

- DVs
  - Number of moves
  - Time on task
  - Number of illegal moves

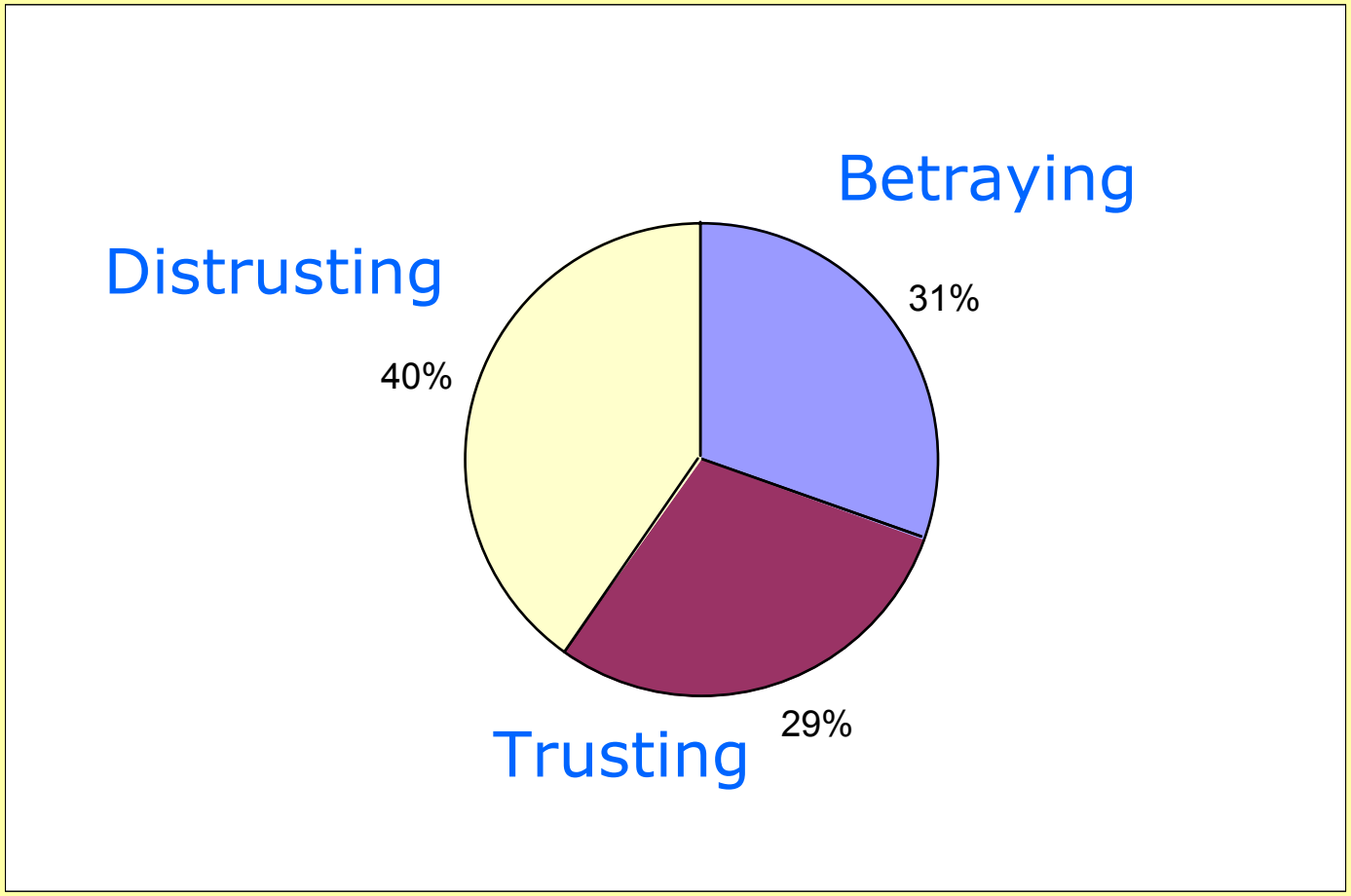
DVs and Analyses

- Repeated measures t-tests on performance on mental rotation
- Individual is the unit of analysis

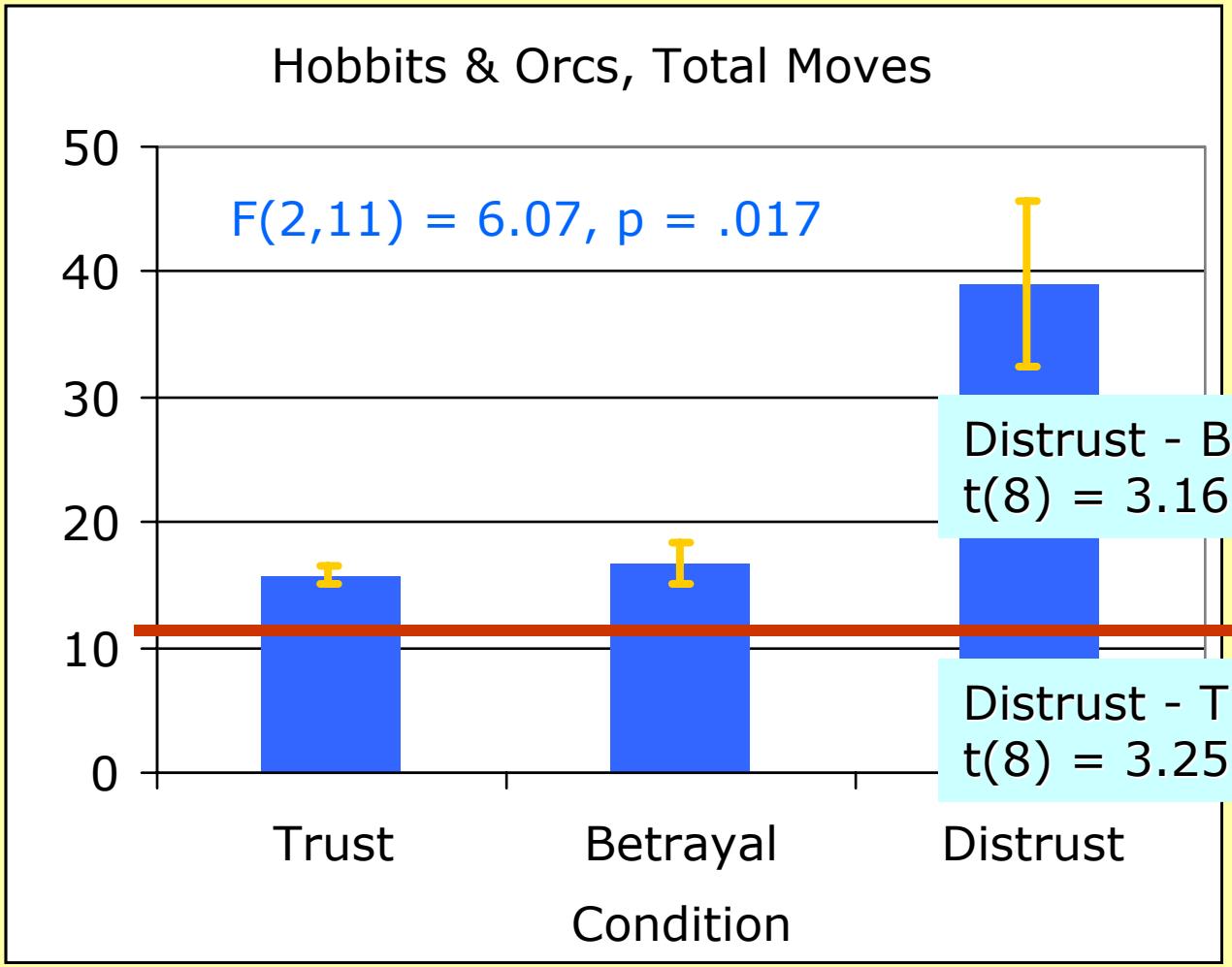
- DVs
  - Percent correct
  - Response time

Results

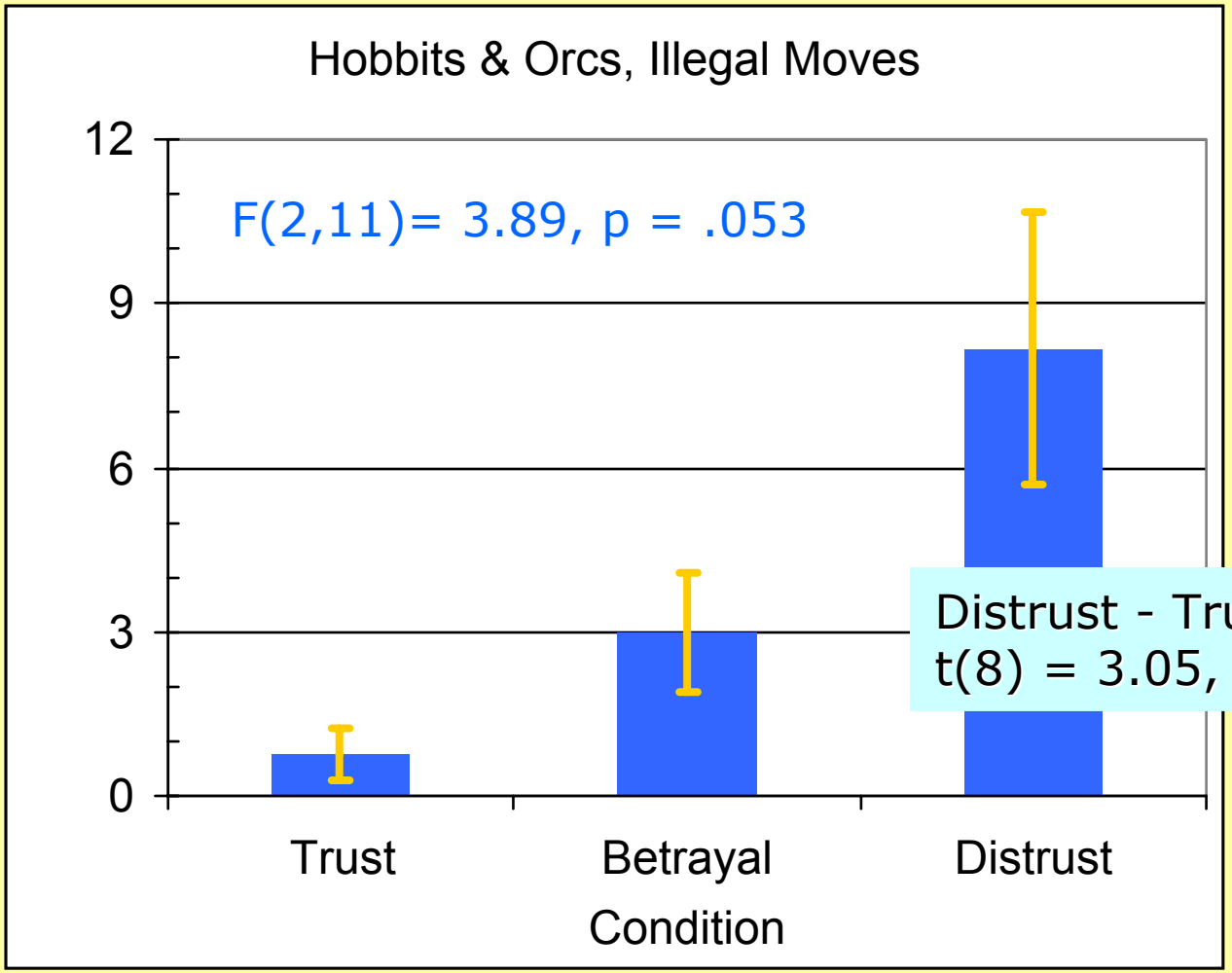
% Teams by trust condition



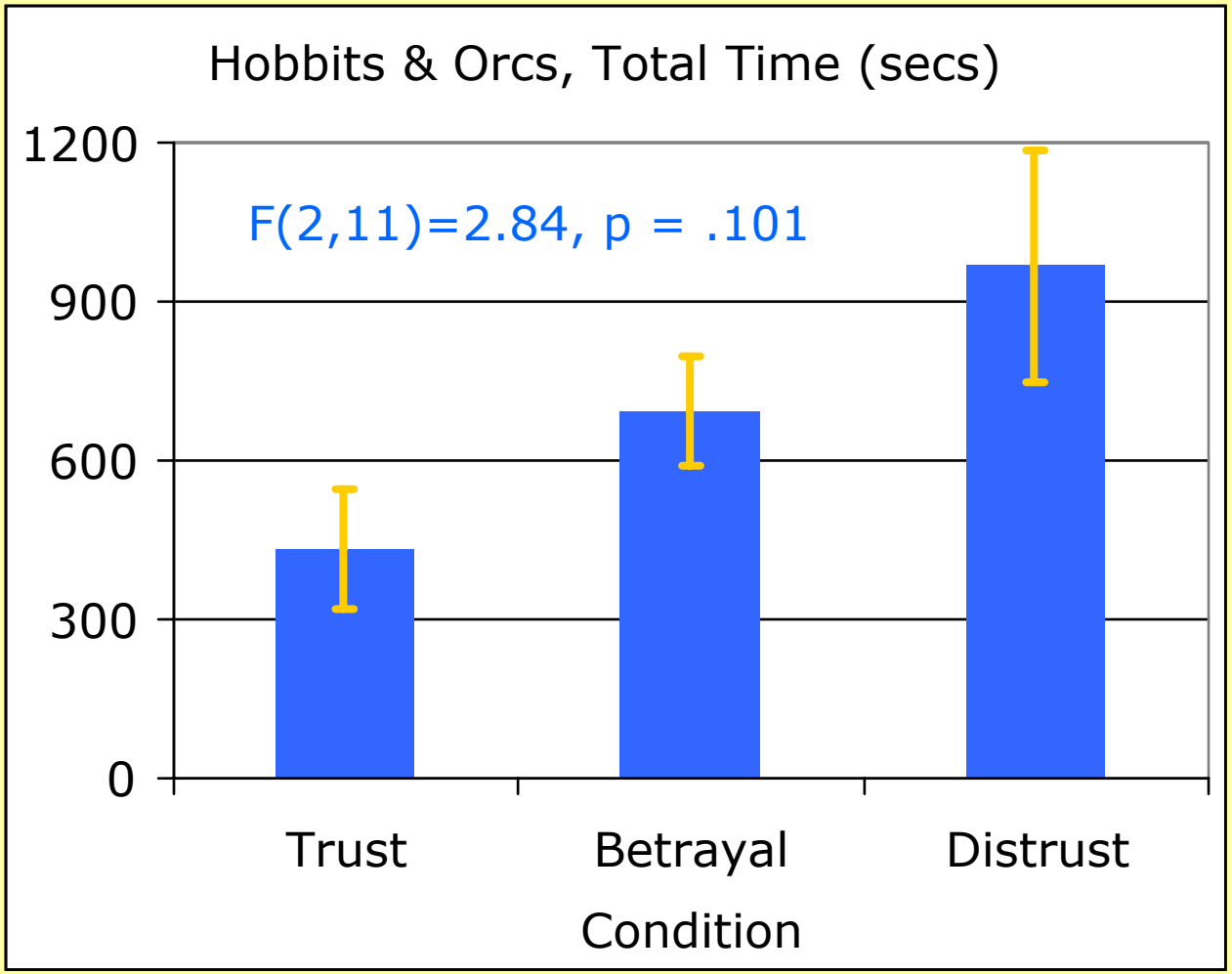
Team problem solving: Total moves, Hobbits and Orcs



Team problem solving: Illegal moves, Hobbits and Orcs



Team problem solving: Time on task, Hobbits and Orcs



Team problem solving: Implications

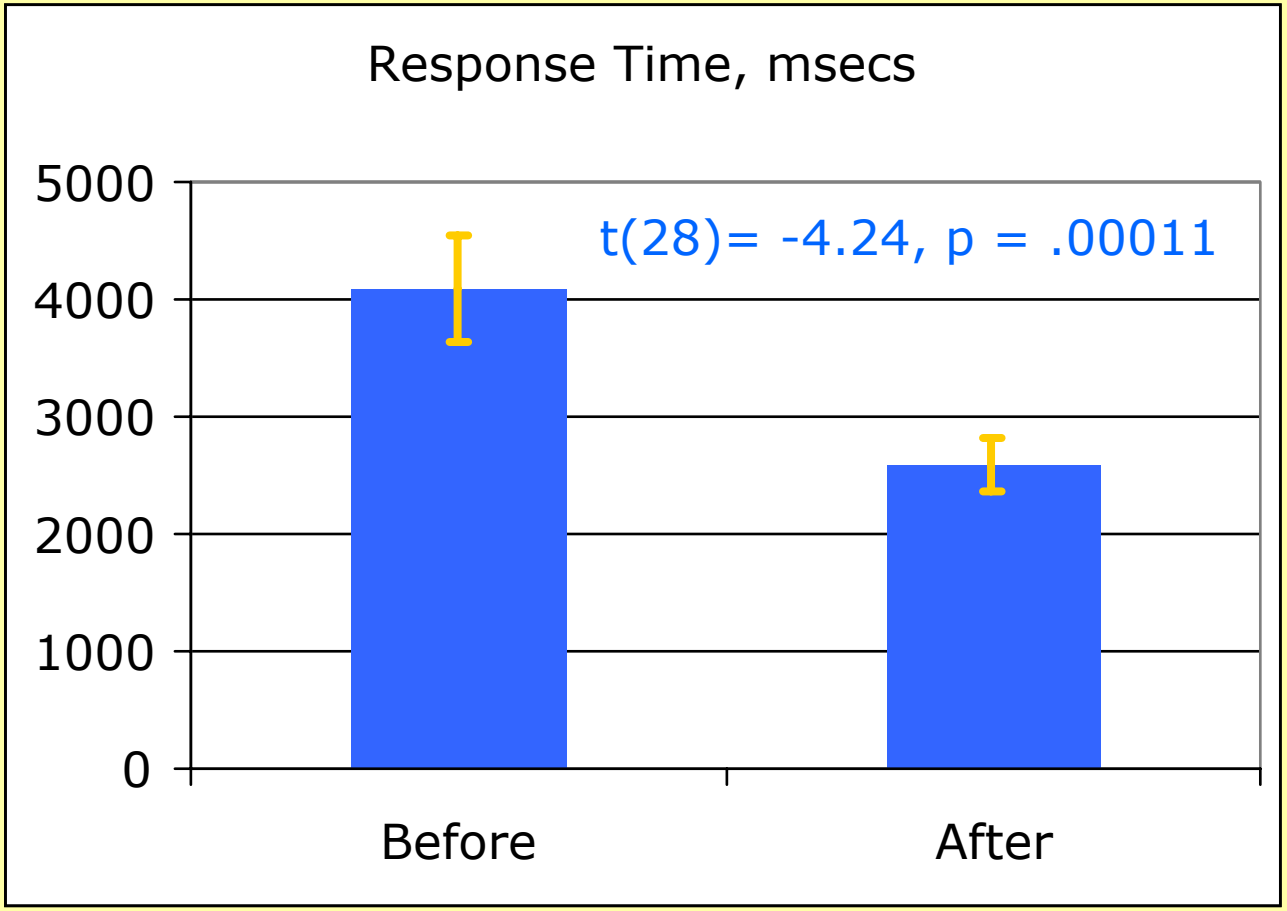
- Levels of trust have a significant effect on team performance

- Results support the hypothesis that team performance in the trusting/trustworthy condition is best overall

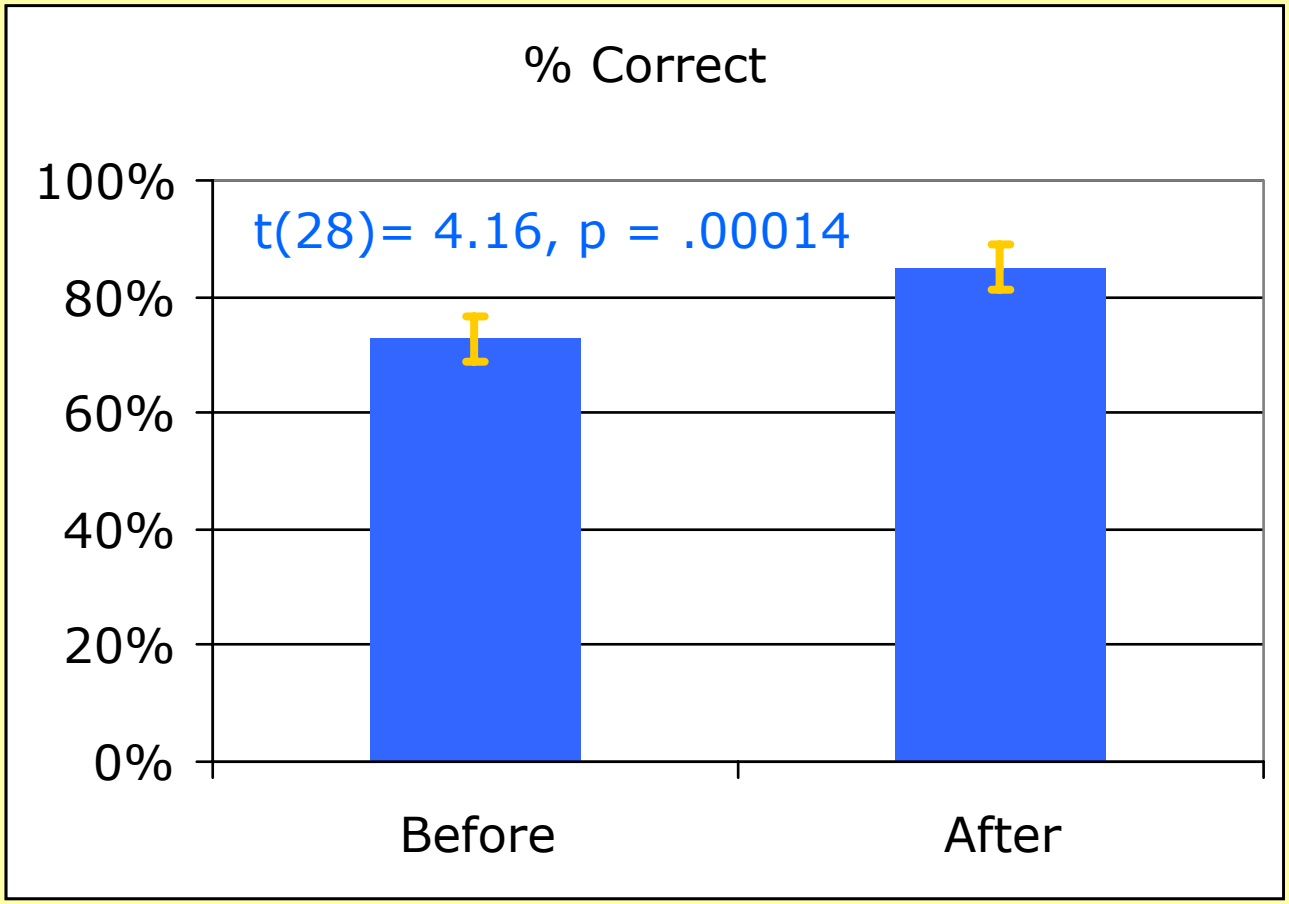
Team problem solving: Implications

- Results do not support the hypothesis that team performance in the betrayal condition is worst
  - Due to dominant betrayer (?)
  - If it were due to dominant betrayer, this result should disappear in a no-talk condition

Individual cognitive processes: Mental rotation



Individual cognitive processes: % correct, Mental rotation

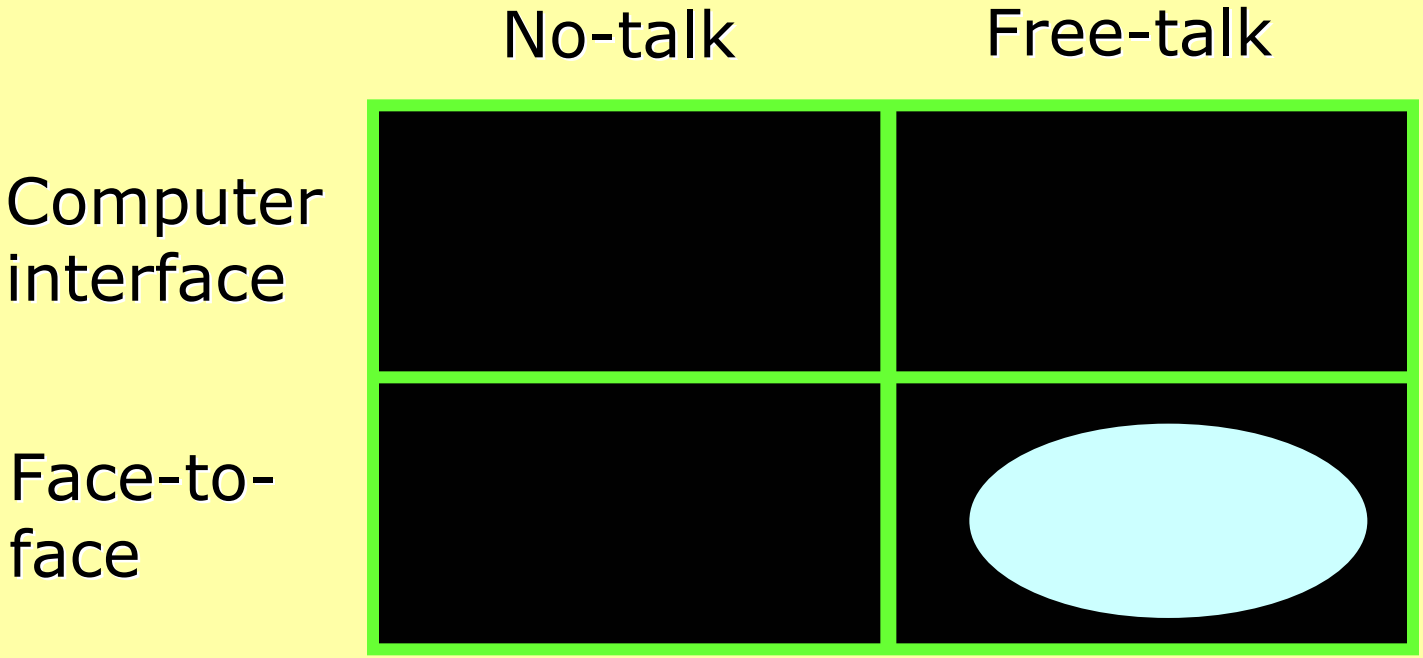


Mental Rotation: Implications

- Significant t-tests on both measures reveal practice effects
  - Faster and more accurate
- There is no negative impact of levels of trust on a basic individual cognitive process

On-going work

- This research is one of four conditions to be tested in a two-by-two factorial design



Acknowledgement

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

The Impact of an Economic Manipulation of Trust on Team Problem Solving

Ben Swan, Andrea Shane, Elizabeth Ratcliff, Nicole Pruitt, Andrew Kearney, Steven Freund, Nathan Fosse, & Kip Smith
Kansas State University

Question and Hypotheses

What effect, if any, do explicit statements of trust or distrust or betrayal have on subsequent team performance?

McCabe, Rassenti, & Smith (1996) have used the extensive-form trust game as a dependent variable to test the predictions of theory and to study theory of mind. This experiment uses it as a quasi-independent variable. Pairs of subjects self-select into one of three levels of trust: Distrusting, Trusting/trustworthy, Betraying. Does this self-selection differentially influence their subsequent performance in the Hobbits and Orcs river crossing problem (Greeno, 1974)?

We expect the level of trust to have a differential impact on team performance but not on individual cognitive processes. Specifically, we expect trusting/trustworthy teams to perform the best and teams that committed or experienced distrust to perform the worst. In contrast, we expect the level of trust to have no effect on repeated play of task involving only individual cognitive processes, e.g., the mental rotation task (Shepard & Metzler, 1971).

Motivation

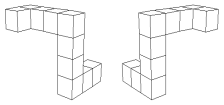
The infantry is changing.

Platoons will no longer train together. The comranderie of boot camp and the front line will be gone and with it the traditional social basis for the establishment of trust. In their place are Objective Force Warriors, the Army of One. These autonomous soldiers will have to cooperate with unknown others like them.

The Army needs to know how trust and autonomy interact to influence teamwork. These experiments are part of a series investigating the interaction of autonomy (leader presence), communication mode, and levels of trust on individual and team performance.

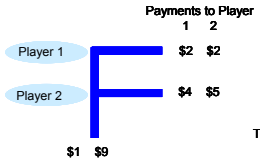
Method

Task 1 and 4:
Mental Rotation
(from Shepard & Metzler, 1971)

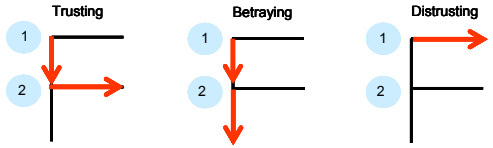


Are these figures the same or different?

Task 2: The Trust Game
Quasi-Manipulation of Trust
(McCabe, Rassenti, & Smith, 1996)



How Players Self-select into Three Levels of Trust:



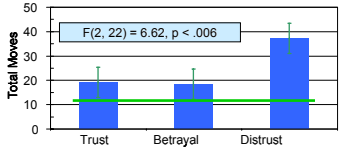
Task 3: Hobbits and Orcs River-Crossing Problem (from Greeno, 1974)



This work is sponsored by Army Research Office
MURI Grant DAAD19-01-1-0621

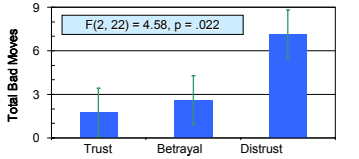
Results, Hobbits and Orcs

Distrusting teams needed more moves



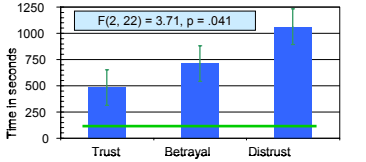
Post- hoc, Total moves
Distrust - Trust: t(13) = 2.68, p = .019
Distrust - Betrayal: t(15) = 2.85, p = .012

Distrusting teams made more bad moves



Post- hoc, Bad moves
Distrust - Trust: t(13) = 2.39, p = .033
Distrust - Betrayal: t(15) = 2.10, p = .053

Distrusting teams took longer to complete the task



Post- hoc, Total time
Distrust - Trust: t(13) = 3.01, p = .010
Distrust - Betrayal: t(15) = 1.85, p = .084

The quality of teamwork differed across the levels of the quasi-manipulation of trust. Distrusting teams worked less efficiently than trusting and betraying teams.

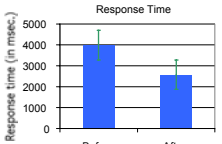
We interpret this result to reflect the asymmetry of information that is unique to the distrust condition.

Results, Mental Rotation

As expected, participants were faster and more accurate at the mental rotation task during the repeated trial.

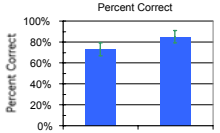
There were no significant differences in the improvement across trust conditions.

Response Time



t(47) = -5.22, p < .001

Percent Correct



t(47) = 5.42, p < .001

Implications

Teams who self-selected into the distrust condition performed less well on all measures of the cooperative problem solving task. This suggests that the quasi-manipulation of trust had a strong impact on the ability of the pair to function as a team.

In contrast, there is evidence of learning in the individual cognitive task. The intervening quasi-manipulation of trust appears to have had little or no effect on spatial reasoning, a basic cognitive process.

The extensive form game appears to be a viable tool for manipulating trust.

Kansas State University
Prof. Kip Smith

Late in 2001 we received permission from UCF to collect data with human subjects. Accordingly, early in the first quarter of 2002, we (1) began to collect data on one of our team problem-solving task, and (2) made heady progress in the coding of a computer simulation of a collaborative task. These accomplishments are discussed in turn.

(1) Dyad problem-solving following a quasi-manipulation of trust

In the first quarter of 2002, we completed the design and software required to collect and analyze data for the first of fours conditions of a study of dyad (two-person team) performance under three different levels of trust (trust, distrust, betrayal). Running the experiment was somewhat complex as we had to insure participant anonymity up to the time they met, half-way through the experiment.

During the course of the experiment, the two participants were asked to do four tasks. The first task was mental rotation. They individually looked at pairs of computer-generated 3-dimensional figures and determined if the two figures were the same or different (mirror images). This task is a classic test of spatial cognition. Our hypothesis is that the levels of trust would have no effect on performance of this individual cognitive ability.

The second task was the extensive form game. This game is taken from experimental economics where it has been used to study non-normative, trusting behavior between strangers. We have turned this traditional use of the game on its head. We use it as a quasi-manipulation of trust. In this quasi-manipulation, the two participants play the extensive form game once. Their play sorted them into one of three categories of trust: distrusting, trusting/trustworthy, and betraying. Thus the subjects themselves rather than the experimenters defined the condition of trust in which they completed the rest of the experiment. The pair received their payoffs and then met face-to-face for the third task.

The third task was a problem-solving game, the classic Hobbits and Orcs river-crossing game. In this task, the team was asked to find a series of legal plays that move game pieces from a specified beginning state to a specified goal state. Our hypothesis is that the levels of trust would have a profound effect on performance of team problem solving. We propose that teams that trusted each other in the economic game would perform better than teams that distrusted each other and than teams that experienced betrayal.

Following the Hobbits and Orcs game, the two participants individually completed a second round of the mental rotation. The repeated play allowed us to test whether the quasi-manipulation of trust had an effect on individual performance and to test for a dissociation in the effect of levels of trust between individual and team performance.

The following is an abstract of a presentation on these data to be presented in May at the Oklahoma/Kansas Judgment and Decision Making Society meeting.

Preliminary studies of the impact of an economic manipulation of trust on team problem solving

The "Hobbits and Orcs" problem is a recent classic in the literature on human problem solving (Greeno, 1974; Jeffries, et al., 1977). We have turned it into a two-person team problem-solving task in order to examine the impact of economic manipulations of trust on team problem solving. The manipulation of trust is a simple version of the extensive-form game, a sequential-play game with monetary payoffs akin to the more familiar, but concurrent-play, game known as the Prisoners' Dilemma (McCabe, Rassenti, & Smith, 1996; McCabe, Smith, & LePore, 2000). In our study, pairs of mutually anonymous subjects first play the extensive-form game. Their play sorts them into one of three categories of trust: distrusting, trusting/trustworthy, and betraying. The pair receives their payoff and then meets face-to-face to solve the Hobbits and Orcs problem. Outcome measures are the number of moves, time to solve the problem, and the number of illegal moves. Process measures include the ratio of cooperative/(cooperative + unilateral) verbal exchanges. Initial results suggest that teams who sort themselves into the trusting/trustworthy condition solve the problem more quickly, more efficiently, and with less discord than teams that were distrusting or betraying.

The results are statistically significant and in accord with our hypotheses.

(2) Dyad cooperation following a quasi-manipulation of trust

We have nearly completed development of software to simulate the navigation of a tugboat in a narrow twisting channel. The task requires cooperation between a pilot (steering) and an engineer (velocity) to avoid crashing the boat into the river bank. The platform allows experiments to control (a) the delay in the feedback between commands and the boat's response, (b) the sinuosity of the channel, and (c) the level of reward for successfully reaching the dock. We plan to use this platform to measure the effects of trust on cooperative task performance.

Kansas State University
Prof. Kip Smith

Second quarter 2002 progress report

In the second quarter of 2002, we (1) continued to collect data on our team problem-solving task, and (2) completed conversion of all tasks to computer-based simulation. In addition, (3) Lauren Murphy completed her MS thesis on the influence of task demand and workload on the occurrence of operational errors in air traffic control. These accomplishments are discussed in turn.

(1) Dyad problem-solving following a quasi-manipulation of trust

In the second quarter of 2002, we completed the collection and analysis of data for the first of four conditions of a study of dyad (two-person team) performance under three different levels of trust (trust, distrust, betrayal). This experiment was detailed in the first quarter progress report.

We made the first presentation of results at a professional meeting, the Oklahoma/Kansas Judgment and Decision Making Society meeting. The abstract of the presentation was provided in the first quarter progress report.

The following is a draft of a paper on these data submitted for presentation at the 2002 annual meeting of the Judgment and Decision Making Society.

Abstract

We use the extensive-form trust game and the classic “Hobbits and Orcs” problem to examine the impact of an economic manipulation of trust on team problem solving. Anonymous pairs of subjects first play the trust game. Their play sorts them into one of three conditions: distrusting, trusting/ trustworthy, and betraying. The pair receives their payoff and then meets face-to-face to solve the Hobbits and Orcs problem. Outcome measures are the number of moves, illegal moves, and time to solve the problem. Teams in the trusting/trustworthy condition solve the problem more quickly and efficiently than teams that experienced distrust or betrayal.

Introduction

Although many studies have been conducted dealing with the effects of trust on group and individual performance (e.g., Gambetta, 1988), few have allowed the subjects to self-select themselves into different categories of trust (e.g., Berg, Dickhaut, & McCabe, 1995). Instead, previous research investigating the effects of trust on performance has placed subjects into groups in which they assumed a trusting or non-trusting relationship with another individual (e.g., Dirks, 1999). In this experiment we use the extensive-form trust game (McCabe, Smith, & LePore, 2000; McCabe, Rassenti, & Smith, 1996) to allow subject pairs to self-select into one of three categories of trust: trusting, betraying, and distrusting.

Measures of subject performance in both individual and cooperative tasks serve as the dependent variables in the experiment. The Hobbits and Orcs river-crossing game (Greeno, 1974; Jeffries, Polson, & Razran, 1977) is used as the test of cooperative performance. The three categories of trust, determined by the trust game, will yield differences in performance levels in the cooperative task. The dyads self-selected into the trusting category are expected to perform at a superior level to dyads in the betraying and distrusting categories. The distrusting dyads are expected to perform at an intermediate level, and the betraying groups at the lowest level.

The test of individual performance is a modified version of Shepard and Metzler's (1971) mental rotation task. This test is presented to the subjects twice: first before they are self-selected into a trust category and also at the end of the experiment. However, it is hypothesized that the outcome of the trust game will not have a negative effect on individual cognitive performance. Due to previous exposure to the test, practice effects should occur and therefore, performance should improve in the second trial. Put simply, a dissociation in the effect of the levels of trust in individual and cooperative performance is expected.

Method

Players

Participants in the study are 28 Kansas State University students enrolled in the general psychology course. Students receive course credit for their participation.

Procedure

Pre-Experimental Protocol

Due to the need to maintain anonymity between the two participants prior to and during play of the trust game, the experiment is presented as two separate studies on the participant sign-up sheets. Additionally, the subjects are instructed to meet the experimenters in different rooms in order to avoid interaction before the experiment. The participants are escorted to separate laboratories where the individual performance and economic game portions of the experiment are completed.

Pretest of Individual Cognitive Performance

The first of four tasks of the experiment involves an individual cognitive test in which subjects perform a mental rotation task. One hundred pairs of three-dimensional figures are presented on a computer screen. The subjects are asked to determine if the pairs of figures are either the same (geometrically equivalent) or different (mirror images) by using the computer keyboard. This task, which is a classic test of spatial cognition, is administered both before and after the economic game.

Quasi-manipulation of trust

The two participants remain in separate laboratories while participating in a trust game derived from experimental economics. The game has two roles, Player 1 and Player 2; each subject takes one of the roles. Figure 1 shows the structure of the game. Three monetary outcomes are possible and each is displayed on a tail of an "F" shaped figure. Player 1, represented by the "1" in the box in the upper left of the "F", is presented with two options. He or she may choose to move right on the top tail of the "F" giving him or herself two dollars, giving Player 2 two dollars, and ending the game. Player 1's second option is moving down toward the middle of the "F" passing control of the game to Player 2. At this point Player 2, represented by the two in the box in the middle of the "F", has two options as well. He or she may chose to move right giving him or herself five dollars, giving four dollars to Player 1, and ending the game. Player 2 also has the option of moving down and receiving nine dollars and giving only one dollar to Player 1.

Each move leading to a specific monetary outcome represents one of three trust categories. A distrusting condition is created if Player 1 decides to move right giving two dollars to both players. If Player 1 makes this move, Player 2 does not receive the opportunity to participate in the game making Player 1's distrust explicit. If Player 1 decides to pass play to Player 2, he or she is considered trusting. Player 2 is considered trustworthy if he or she moves right giving him or herself five dollars and giving Player 2 four dollars. A betraying condition is created if Player 1 has trusted Player 2, and Player 2 moves down and receives nine dollars and gives only one dollar to Player 1. Thus, subject dyads are self-selected into one of three categories: trusting, distrusting, or betraying. Additionally, if Player 1 is distrusting, Player 2 does not have the opportunity to play and therefore, his or her propensity to be trustworthy or betraying is never revealed.

After the subjects complete the trust game, they are brought together into one laboratory and are presented with the monetary award in cash based upon the outcome of the trust game.

Cooperative problem solving task

The third task in the experiment is a test of cooperative group performance. The classic Hobbits and Orcs river-crossing game is utilized to test the effects of the three levels of trust on group performance. Using a game board with a river running through the middle, the subjects are directed to move three Hobbits and three Orcs across the river using a small boat. Three rules exist concerning the manner in which the Hobbits and Orcs may cross the river: 1) the Orcs may not outnumber the Hobbits on either side of the river at any time, 2) only two characters are allowed in the boat at a time, and 3) the subjects are allowed 100 moves or 45 minutes to complete the activity. The subjects' participation in the game is videotaped in order for the experimenters to accurately collect the data from the subjects' performance.

Posttest of Individual Cognitive Performance

The mental rotation task that was presented at the beginning of the experiment is repeated by each subject as the last portion of the experiment. This task is completed in separate laboratories as during the first trial.

Results

Individual cognitive performance

Two dependent variables were utilized to measure individual performance in the mental rotation task: response time and percent correct (Figure 2). Results show that the mean response time decreased in the posttest, $t(28)=-4.24$, $p=0.00011$, which is consistent with the hypothesis. A practice effect is also apparent for the percent correct data, $t(28)=4.16$, $p=0.00014$. Consistent with the hypothesized results, performance on the mental rotation task improved during second trial. It appears as though neither the trust condition from the economic game, nor the interaction during the cooperative problem-solving task had adverse effects on individual performance.

Cooperative problem solving

Prior to the experiment, a pretest using 54 undergraduates in an upper-level psychology course was conducted. Seventy-two combinations of the trust game were presented in the pretest to determine monetary combinations that would place approximately one-third of the subjects in each of the three trust conditions. Due to the fact that the dyads self-select themselves into the trust conditions, it was necessary to find a combination of monetary awards that would yield approximately equal numbers of subjects in each condition. The combination used for the experiment is shown in Figure 1. The results of the trust game placed 6 (42%) of the dyads in the distrusting condition, 4 (29%) in the betraying condition, and 4 (29%) in the trusting condition (Figure 3).

The effects of the trust conditions on group performance, represented by the Hobbits and Orcs cooperative task, were analyzed using three dependent variables: total number of moves, number of illegal moves, and total time elapsed during the task. These data are presented in Figures 4, 5, and 6, respectively. In regard to the total number of moves, an ANOVA found significant results, $F(2,11)=6.07$, $p=0.017$. A post-hoc test revealed that the performance of the betraying group was superior to that of the distrusting group, $t(8)=3.16$, $p=0.013$, and that the performance of the trusting group was superior to the distrusting group as well, $t(8) = 3.25$, $p=0.011$, Figure 4. These statistics contradict the hypothesis that stated the order of performance from best to worst to be: trusting, distrusting, betraying.

The data concerning the number of illegal moves within each condition, Figure 5, also exhibited significant results, $F(2,11)=3.89$, $p=0.053$, with a post-hoc test signifying the distrusting group committing more illegal moves than the trusting group, $t(8)=3.05$,

p=0.015. Results involving the total time elapsed, Figure 6, during the cooperative task did not yield significant results.

Discussion

Experimental manipulation

The use of self-selection as a means to induce differences within group performance in the economic trust game appears to be viable experimental paradigm for investigating the impact of trust on team performance. Not only did subjects self-select themselves into conditions so that each condition was represented at nearly an equal level, but also, significant results in regard to cooperative performance were found.

Cognitive performance

Neither the trust game nor the cooperative problem-solving task had an impact on individual performance. Performance on the posttest of the mental rotation task was superior to that of the pretest implying that learning effects took place. As expected, the dyad trust level did not impede this basic cognitive process.

Cooperative problem-solving hypotheses

The experimenters predicted that the performance level of the trust groups from best to worst would be: trusting, distrusting, betraying. This hypothesis was not supported by the post-hoc tests. In regard to the total number of moves, the performance level from best to worst was: trusting, betraying, distrusting. In regard to the number of illegal moves, the trusting groups performed at a superior level to the distrusting group; however there was no significant difference between the performance of the trusting group and betraying group or between the performance of the betraying group and the distrusting group. A review of the videos of the subjects' participation during the Hobbits and Orcs game showed that the betrayers assumed a dominant role during the game. In fact, the dominant betrayer often took complete control of game play. This finding may account for the unexpected results.

If this hypothesis regarding a dominant betrayer is true, the effectiveness of the betraying group's performance will reduce if the opportunity for dominance during the game is eliminated.

Future Work

The proposed experiment represents one of four conditions in a 2 X 2 factorial design. The two factors of the design are communication mode and degree of interaction. The levels that comprise the communication mode are free talk and no talk. The two levels of the degree of interaction are anonymous (in different rooms and

having no interaction except for game play) and familiar (in the same room at adjacent computers). The current experiment represents the familiar, free talk condition.

The trust game and the Hobbits and Orcs cooperative problem-solving game have been computerized; future experimental procedure will use the computerized versions of both games. The current condition (familiar, free talk) will be repeated in order to test for differences due to presentation or procedure.

Utilizing the computer interface will make it possible to guarantee anonymity. In the anonymous conditions, the subjects will not see each other nor will they have the opportunity to interpret non-verbal cues. The computer interface will also intensify differences due to the three levels of trust since the subjects will not have opportunity for direct interaction. Additionally, in the free talk condition, the two subjects will communicate via a web-based chatroom.

The no-talk level is expected to eliminate the opportunity for a dominant player to take total control of the cooperative problem-solving game. If the hypothesis stating that the superiority of the betraying group's performance to that of the distrusting group is due to a dominant betrayer is correct, the no-talk conditions will verify this hypothesis.

References

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. Games and Economic Behavior **10**, 122-142.

Dirks, K. (1999). The effects of interpersonal trust on work group performance. Journal of Applied Psychology **84**, 445-455.

Gambetta, D. (1988). Trust. New York: Basil Blackwell.

Greeno, J. (1974). Hobbits and Orcs: Acquisition of a sequential concept. Cognitive Psychology **6**, 270-292.

Jeffries, R., Polson, P., & Razan, L. (1977). A process model for missionaries-cannibals and other river crossing problems. Cognitive Psychology **9**, 412-440.

McCabe, K., Rassenti, S., & Smith, V. (1996). Game theory and reciprocity in some extensive form experimental games. Proceedings of the National Academy of Science **93**, 13421-13428.

McCabe, K., Smith, V., & LePore, M. (2000). Intentionality detection and "mindreading": Why does game form matter? Proceedings of the National Academy of Science **97**, 4404-4409.

Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. Science **171**, 701-703.

Figures

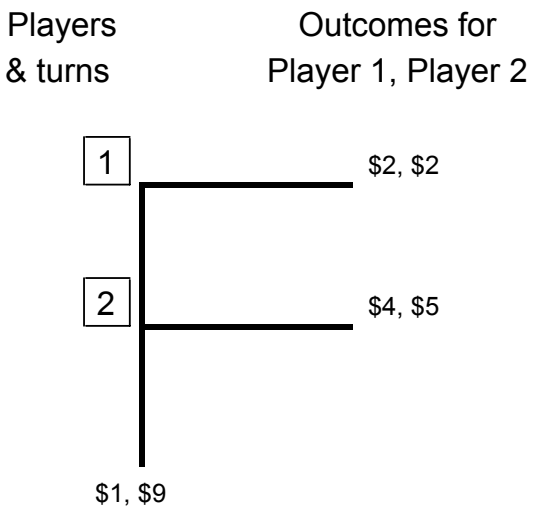


Figure 1. The extensive-form trust game used as the stimulus in the quasi-manipulation of trust. The numbers in the squares represent Player 1 and Player 2. Player 1 has the first move. A move to the right is a distrusting move that nets both players $2. A move down is a trusting move that passes play to Player 2. A move right by Player 2 is a trustworthy move that nets Player 1 $4 and Player 2 $5. A move down by Player 2 betrays Player 1's trust.

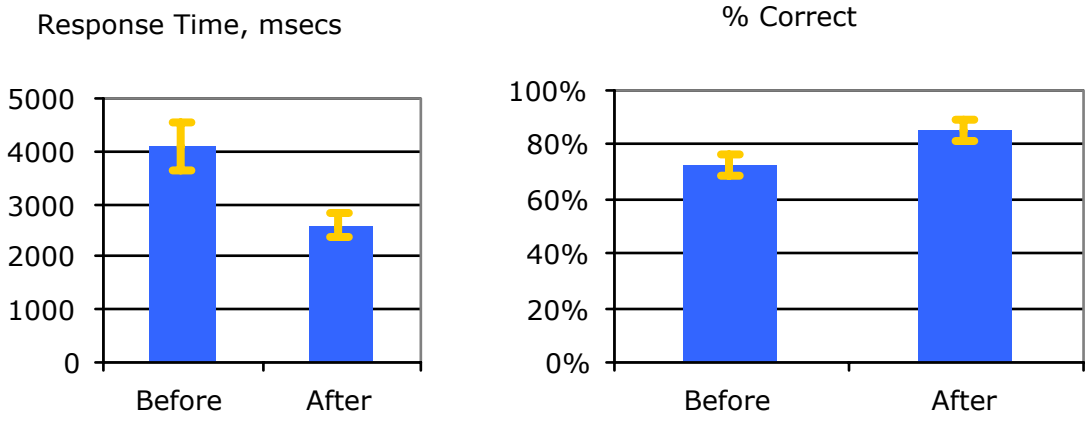


Figure 2. Response times and percents correct for the two trials of the mental rotation task. As expected, consistent with learning, times decreased and accuracy increased in the second trial.

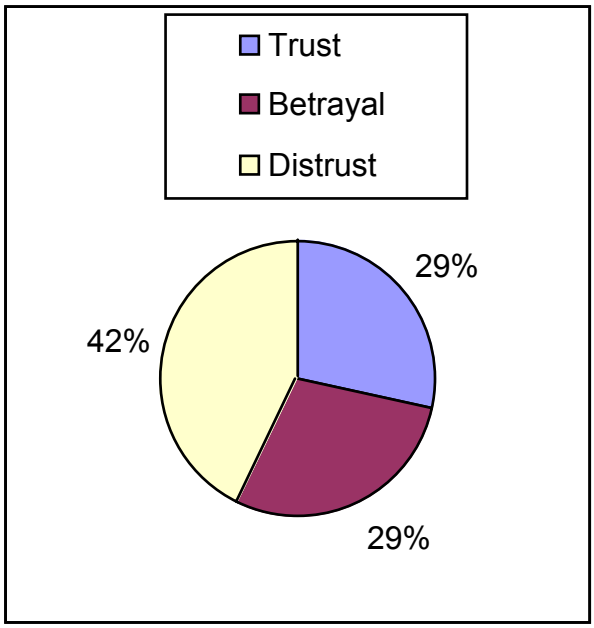


Figure 3. The approximately even distribution of dyads per trust category from the quasi-manipulation of trust.

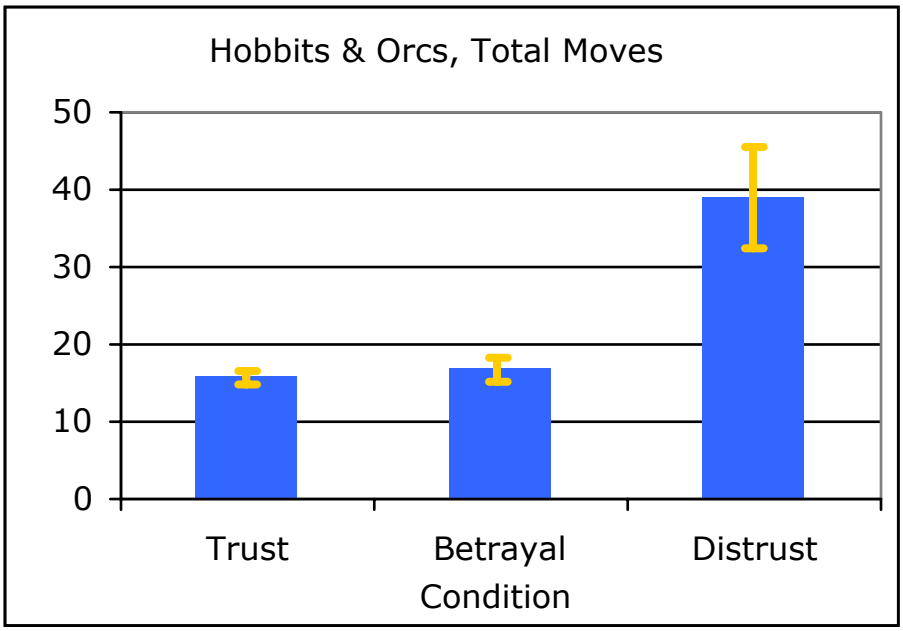


Figure 4. Mean number of total moves in the cooperative problem solving task by dyads in the three categories of trust.

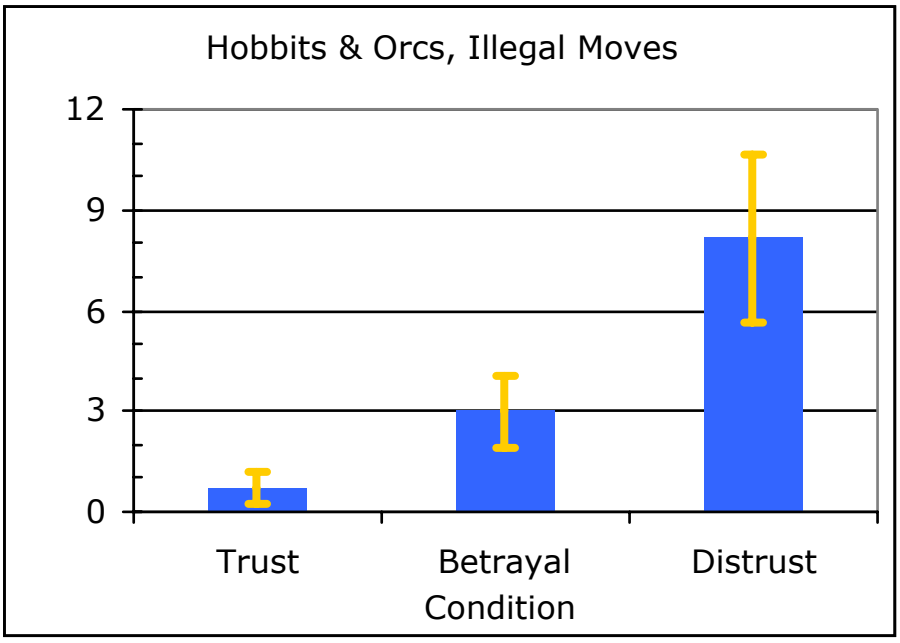


Figure 5. Mean number of illegal moves in the cooperative problem solving task by dyads in the three categories of trust.

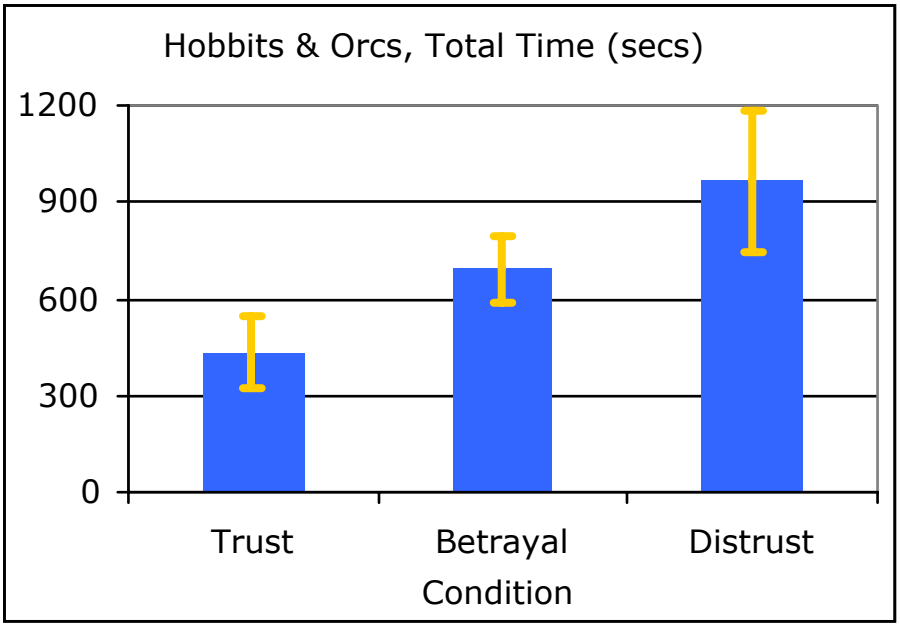


Figure 6. Mean time to complete the cooperative problem solving task by dyads in the three categories of trust.

(2) Computerization of tasks

We have completed conversion of all experimental stimuli to PCs. These stimuli are (A) the extensive-form trust game, (B) the hobbits and Orcs problem, (C) the boat game, and (D) the mental rotation task. These are discussed in turn.

2A) The extensive-form trust game

We use a simple version of the extensive-form trust game as the stimuli for a quasi-manipulation of trust. The extensive-form game is a sequential-play two-player game with monetary payoffs akin to the more familiar, but concurrent-play, game known as the Prisoners' Dilemma (McCabe, Rassenti, & V. Smith, 1996; McCabe, V. Smith, & LePore, 2000). Play in the game sorts pair of participants into one of three categories of trust: distrusting, trusting/ trustworthy, and betraying.

The game and its instructions are now available on networked Wintel computers. Two players, Player 1 and Player 2, sit in different laboratory rooms and communicate solely through their play in the game. They receive their payouts in cash.

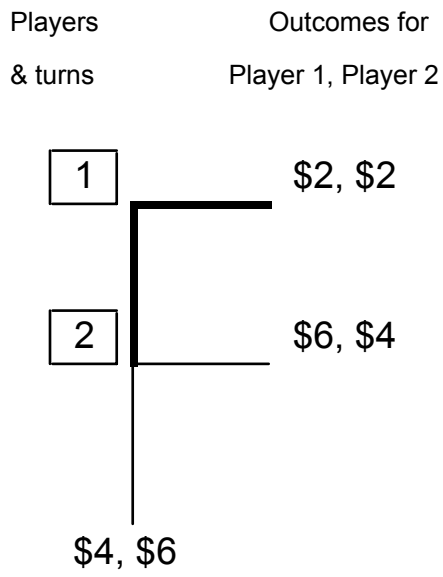
The simplest way to understand the game is to follow the instructions to players for Player 1. A corresponding set of instructions are provided to Player 2.

The Extensive Form Game
Instructions for Player 1

The purpose of this research is to study choice behavior between monetary alternatives. This project is being conducted by Professor Kip Smith in the Psychology Department at Kansas State University. You will an amount of cash for participating. The amount you earn depends on your performance in this experiment.

In this experiment, you will be working with another participant, who is located in another room. For this experiment, the other participant is called your counterpart. You are called "Player 1." Your counterpart is called "Player 2." Both you and your counterpart have complete access to all information.

The figure below is an example of the game. You will play it once and only once. The monetary outcomes in this example are not identical to those in the game you will play. The purpose of the example is to explain how to play the game.

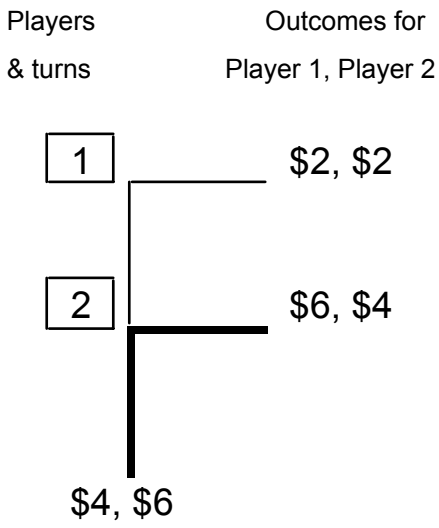


You, as Player 1, are represented by the "1" in the box in the upper left corner of the figure. Your counterpart, Player 2, is represented by the "2" in the box in the middle of the left side of the figure.

Player 1 has the first move. You have two possible moves from the "1" in the box: either to the right or down. These moves are represented by the bold, thick lines in the figure above.

If you chose to move right, you will receive $2 and Player 2 will receive $2. At this point the game will end without Player 2 getting a chance to play.

If you chose to move down, you pass play to Player 2 and it becomes Player 2's turn.



Player 2 has two possible moves from the "2" in the box: either to the right or down. These moves are represented by the bold, thick lines in the figure above. If Player 2 chooses to move right, you will receive $6 in cash and your counterpart will receive $4 in cash. At this point the game will end.

If Player 2 chooses to move down, you will receive $4 in cash and your counterpart will receive $6 in cash. At this point, the game will end.

It is important to remember that once you or Player 2 choose a monetary outcome, the game is over. You will play the game once and only once.

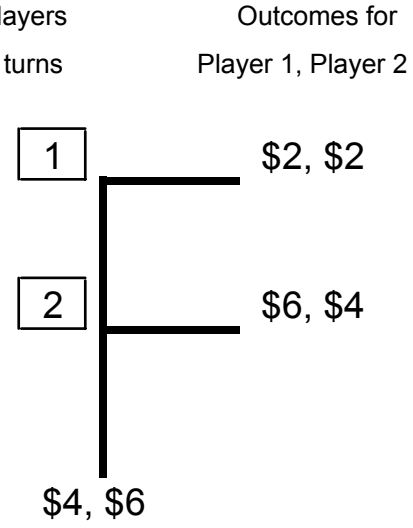
Please notify the experimenter now if you have any questions about how to play the game.

Pretest – Player 1

Here is a sample game. Please indicate to the experimenter the two moves you can make. Remember, you are Player 1.

What happens if you make choose the first move?

What happens if you choose the second move?



Any questions?

<end of instructions to players>

Teams that end the game by having Player 1 move right (along the top arm of the "F") experience distrust: Player 1 does not trust Player 2 to move right along the lower arm of the "F". These teams sort themselves into the distrusting condition.

If Player 1 moves down, passing play to Player 2, Player 1 is making an explicit statement of trust in Player 2.

Teams that end the game by having Player 2 move right (along the lower arm of the "F") experience both trust and trustworthiness: Player 2 reciprocates Player 1's statement of trust with a move that sacrifices self-interest and nets the team the greatest good. These teams sort themselves into the trusting condition.

Teams that end the game by having Player 2 move down (to the bottom of the "F") experience trust and betrayal: Player 2 betrays Player 1's statement of trust with a move motivated by self-interest. These teams sort themselves into the betraying condition.

Thus, the teams sort themselves into one of three conditions. This is the basis for our quasi-manipulation of trust.

2B) The hobbits and Orcs problem

The "Hobbits and Orcs" problem is a recent classic in the literature on human problem solving (Greeno, 1974; Jeffries, et al., 1977). We have turned it into a two-person team problem-solving task in order to examine the impact of an economic manipulation of trust on team problem solving.

The game and its instructions are now available on networked Wintel computers. The game is a distributed simulation. Two players, Player 1 and Player 2, sit in different laboratory rooms but see the same information on the screen. They take turns issuing instructions.

The screen shows a static woodland scene bisected by a river. Beside the river is a boat, with room for 2 passengers. At the beginning of play, six figures march together to one bank of the river. Three are "Hobbits" and three are "Orcs." Orcs are blood-thirsty but cowardly; hobbit is a particularly delicious meal. Hobbits are kindly creatures, not prone to violence, but will fight tenaciously if cornered. Orcs know this and, being cowardly, will not try to eat the Hobbits if they are evenly matched or outnumbered.

The object of the game is to transport all three Hobbits and all three Orcs to the other side of the river without losing any Hobbits. There are two rules. First, only two "passengers" can fit in the boat at one time. The only other rule is that the Orcs must never outnumber the Hobbits on any one side of the river. (If the Orcs outnumber the Hobbits, the Orcs will eat the Hobbits. This is not good.)

The players take turns putting characters into the boat and launching it. A click on a character prods it to jump into the boat. A click on the launch button sets the ferry in motion to the other bank. Once the boat reaches the far bank, the (one or two)

characters in the boat jump out. The players are allowed to work on this game until either they solve it, make 100 futile moves, or spend 45 minutes trying.

The optimal solution takes 11 moves. The computer program records the moves, their number, and detects and counts "illegal" moves. An illegal move is one that puts more Orcs than Hobbits on one bank of the river. Whenever this happens, the program treats both players to a vision of the Orcs eating the hobbits, complete with gore. The program then reverses the move that produced the feast, gives the players a "take over", and counts the illegal move. Play resumes from the previous legal state.

At the end of successful play, the six figures abandon the boat and march together into the woods on the far bank of the river.

2C) The boat game

The boat game was inspired by a dozen years of watching the masterful crews of Mississippi River pushboats (tugboats that push rather than pull) navigate as many as six yoked barges through the bridges of St Paul and Minneapolis MN with nary a mishap. These barge-pushboat complexes have considerable momentum. Taking a turn requires coordination and planning between the capture who steers the boat and the engineer who controls the engines. The coordination is made complex by the significant time lag between the issuing of an instruction and its realization.

This computer-based game is a distributed simulation of tugboat operation. Two players, Player 1 and Player 2, sit in different laboratory rooms and communicate solely through their play in the game. One player is the captain responsible for steering the boat. The other player is the engineer responsible for velocity.

The goal of the game is to get to finish line as fast as possible. The game has a monetary incentive. At the end of the game, when the players get to the finish line, they split the total of the dollar amounts shown on the "fuel gauge" and the "clock". The "clock" counts down from $15.00 at the rate of $0.01 per second. The faster the game is completed, the more money the players take home.

There are a few complications. First, running into one of the river banks causes the boat to stop for a full five seconds. At the end of its stall, the boat pops back into the center of the river with zero forward velocity. Second, each and every instruction to increase speed reduces the amount of fuel left. The "fuel gauge" starts at $15.00 and is decrimented by $0.05 with each instruction to increase speed. It is very easy to burn a dollar of fuel in an attempt to hydrofoil down straight stretch of river. Doing so invariably makes it impossible to turn the corner without either hitting the bank or slowing down. It is entirely possible to run out of fuel. No one has done so yet, but we shall see.

The student designers of the game would be splitting roughly $18.00 (the sum of the amounts remaining on the fuel gauge and clock) after several months of practice. We expect participants in our experiments to take home considerably less. The

experimental question is “Do the three conditions of trust influence levels of cooperation in the boat game?” And more generally “What is the influence of trust on cooperative behavior?” This experiment will be run in the fall semester, 2002.

2D) Mental rotation

This task was completed in previous quarters and is mentioned here for the sake of completeness. The bitmap images of the objects are presented by an application prepared using SuperLab, commercial experimentation software.

(3) On the influence of task demand and workload on operational errors in air traffic control

The following is the abstract a paper on these data submitted for publication to the journal of the Human Factors and Ergonomic Society.

This paper investigates the relationship between task demand, the operator’s psychophysiologic and behavioral response to that demand, and the occurrence of errors. The domain of investigation is en-route air traffic control (ATC). We controlled the level of task demand (aircraft traffic count) in a series of challenging but realistic en-route air traffic scenarios and monitored participants’ heart rates, communications, and traffic-controlling actions. Traffic count and operational errors due to errors of omission were our behavioral measures of task demand and response to task demand, respectively. Heart-rate variability at 0.1 Hz (the Traube-Hering-Mayer wave) was our psychophysiologic measure of response to task demand. This combination of metrics allowed us to distinguish between task demand and operator response to task demand. Operational errors due to errors of omission were found to be equally likely after a peak in traffic count as they were during a peak. Heart-rate variability reflected the operator’s performance and not the demands of the task environment: it varied monotonically with the occurrence of errors and deviated systematically from the pattern of task demand.

Kansas State University
Prof. Kip Smith

Third quarter 2002 progress report

The MURI-supported research at K State focuses on identifying and mitigating (a) the effects of distrust on team performance and (b) the interaction between levels of trust and modes of communication between team members and leaders. We view distrust as a source of stress that is likely to have a deleterious impact on teamwork and cooperative behavior. We expect anonymity and the lack of leader presence to exacerbate this impact. This basic research on trust, communication, cooperation, and team work addresses issues raised by the prospect of an 'electronic soldier' who is physically and psychologically distant from team members and leaders.

During the third quarter of 2002, we have been conducting two laboratory experiments and preparing a field experiment, and have nearly completed the software development for a fourth experiment. These four endeavors are named below and discussed in turn.

- On the effect of play in an economic trust game on subsequent team problem solving
- On the effect of team performance in a distributed, cooperative task on subsequent play in an economic trust game
- On the effect of leader presence on team performance under fire on a paintball lane
- On the development of trust and cooperation by players of a fantasy game

(1) On the effect of play in an economic trust game on subsequent team problem solving

This experiment investigates the effect of a quasi-manipulation of trust at three levels (trust, distrust, and betrayal) on team work and individual cognitive performance. The four treatment conditions fully cross presence at two levels (face-to-face, anonymous) and communication at two levels (free talk, no talk).

Our second quarter report contained the draft of an article on the first phase of this work. The paper describes that we have found striking evidence for the deleterious effects of distrust in the face-to-face, free talk condition. It also describes a clear dissociation of the effect of levels of trust on individual and team performance. The paper will be presented in November at the 2002 annual meeting of the Society for Judgment and Decision Making. During the third quarter, we continued data collection in all four treatment conditions using the computerized versions of experimental stimuli described in the second quarter progress report.

Table 1 – The four treatment conditions

		Presence	
		Face to face	**Anonymous**
Communication	**Free talk**	Same room, ordinary conversation	Different rooms, radio communication
	No talk	Same room, no conversation allowed	Different rooms, no communication

Synopsis of the experiment – Hobbits and Orcs

In this laboratory experiment, the two participants perform four tasks. The first is a mental rotation task, a classic measure of visuospatial ability. The second task is the extensive form economic trust game shown in Figure 1. The two participants play this game for cash once. The game has three possible outcomes. In the analysis of this experiment, we treat the outcome of the game as a quasi-manipulation of trust at three levels (distrust, trust, and betrayal), that is, as an independent variable. Participants receive the amounts of cash indicated immediately upon completion of the game.

The third task in the experiment is a team problem solving task, the Hobbits and Orcs river-crossing game. The goal of the game is to move three Hobbits and three Orcs across a river using a small boat. Three rules constrain play: 1) the Orcs may not outnumber the Hobbits on either side of the river at any time (or else the Orcs eat the Hobbits, not a good thing), 2) at least one but no more than two characters may be in the boat at a time, and 3) the subjects are given 100 moves or 45 minutes to complete the activity. One player is responsible for launching the boat from the left bank, the other for launching the boat from the right bank. Dependent measures include the number of moves, the number of illegal moves (e.g., allowing a Hobbit to get eaten), and the amount of time taken. We have found that these simple measures are remarkably sensitive to the outcomes of the economic trust game. Trusting pairs perform better than distrusting pairs and betraying pairs.

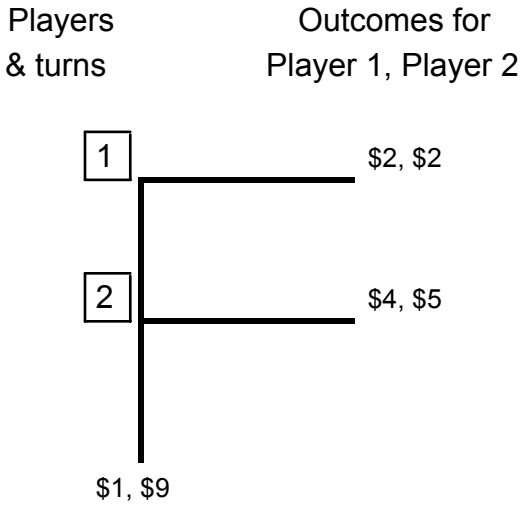


Figure 1. The extensive-form trust game used as the stimulus in the quasi-manipulation of trust. The numbers in the squares represent Player 1 and Player 2. Player 1 has the first move. A move to the right is a distrusting move that nets both players $2. A move down is a trusting move that passes play to Player 2. A move right by Player 2 is a trustworthy move that nets Player 1 $4 and Player 2 $5. A move down by Player 2 betrays Player 1's trust.

The fourth task repeats the first, the mental rotation task. This provides the opportunity to perform a repeated measures analysis of individual performance. As expected, most participants benefit from repeated play, getting both faster and more accurate on the second trial. This improvement appears to be impervious of the outcomes of the economic trust game and the cooperative problem solving game. Players from trusting pairs perform no better than players from distrusting and betraying pairs. Thus, we have documented a strong dissociation in the effects of levels of trust on individual cognitive performance and teamwork.

(2) On the effect of team performance in a distributed, cooperative task on subsequent play in an economic trust game

This laboratory experiment investigates of effect of team work on trust. As in the first experiment, the four treatment conditions fully cross presence at two levels (face-to-face, anonymous) and communication at two levels (free talk, no talk), Table 1. Unlike the first experiment, this experiment uses the extensive form economic trust game, Figure 1, as a dependent variable.

Turning a Mississippi River pushboat (tugboats that push rather than pull) requires coordination and planning between the captain who steers the boat and the engineer who controls its velocity. The coordination is made complex by the significant time lag between the issuing of an instruction and its realization. The heart of this experiment is a computer-based, distributed simulation of pushboat operation. One

player is the captain responsible for steering the boat. The other player is the engineer responsible for velocity. The goal of the game is to get to finish line as fast as possible. The game has a monetary incentive. At the end of the game, when the players get to the finish line, they split the total of the dollar amounts shown on the "fuel gauge" and the "clock". The "clock" counts down from $15.00 at the rate of $0.01 per second. The "fuel gauge" starts at $15.00 and is decrimented by $0.05 with each instruction to increase speed. It is easy to run out of fuel.

At issue in this experiment is the impact of the quality of cooperation (money received) during play of the boat game on subsequent play of the economic trust game. Data collection began in early September in the anonymous, no talk condition. There are too few data at this time to support inferential analysis.

(3) On the effect of leader presence on team performance under fire on a paintball lane

This experiment will be conducted at the local paintball course. The prospect of being hit by paintballs poses a situation that is as stressful and as combat-realistic as University research can hope to provide. The purpose of the experiment is to study the effect of leader presence and communication mode on trust in the leader and between teammates under stressful conditions.

It has taken most of the summer and a month of the fall semester (the entire third quarter) to satisfy and gain the approval of the K State Institutional Review Board's Committee on Human Subjects to conduct this experiment. The first evening of data collection is scheduled for Wednesday September 25.

Teams of two subjects will be asked to complete a paintball assault lane against an "enemy" strongpoint. The lane consists of four pairs of inflated "bunkers" that provide shelter from the enemy's fire. One of the experimenters, a high-ranking ROTC cadet, is the team's "leader." The team's objectives are to follow their leader's orders and to avoid being hit by paintballs. The leader follows a set script to give two types of orders. The first is to retrieve pairs of small flags and to advance to the next bunker. The second is to provide covering fire for their teammate when (s)he advances. The flags are placed in the open at fixed distances from the bunkers. The enemy, a second experimenter who has won sharpshooting competitions, has a clear shot at all the flags.

Two observers, one per subject, record the elapsed time between when an order is given and when the first of the two flags is retrieved. This response time is our chief dependent measure of the subject's level of trust in the leader.

There are two conditions of the lane. In the first condition, the leader is physically present (and under fire) on the lane. This condition simulates current military practice. In the second condition, the leader is nowhere in sight and delivers his orders via radio. This condition simulates the situation that will be faced by the "electronic soldier." The order of conditions will be counterbalanced.

Subjects will also be asked to fill out a questionnaire three times during the experiment, before the experiment begins, between the two conditions, and after the second condition. The questionnaire solicits subjective ratings of experienced anxiety, stress (exertion), and trust in the leader and team member. Other dependent measures include the number of times hit, the number of paintballs used, and the total time the team needs to complete the lane.

(4) On the development of trust and cooperation by players of a fantasy game

Two of the chief criticisms of the validity of laboratory experiments are that they typically last for less than an hour and are not sufficiently involving for the subjects to generalize beyond the laboratory. During the third quarter, three undergraduate research assistants have designed and constructed an experiment that should overcome these concerns. As currently formulated, the two-person game will last for approximately 10 hours over a period of several weeks and will, we hope, seduce the student participants into wanting it to continue.

The experimental platform is Neverwinter Nights (© Atari and Bioware), a 'dungeons and dragons' software game that gives users access to the source code and enables them to script their own microworlds. Our microworld is filled with monsters and magic, vixens and a bard, but is based firmly on Deutsch and Krauss' (1960) trucking game. In this interactive fantasy game, two players are heroes on a mission to save the world. They must transport quantities of elixir between pairs of springs to keep both in balance and the forces of evil at bay. There are two long winding roads and one shorter one-lane road between the springs. If the players meet on the one-lane road, one must turn around and allow the other to proceed. There are two gates on the one-lane road. One player has the key to one of the gates and the other player has the key to the other gate. Our 221st century realization of Deutsch's microworld gives the subjects ample opportunities to collaborate, to negotiate, to distrust each other, and to betray each other.

The game records the dialog boxes the players select, the words they type in, and the moves they make. For example, locking a gate to prevent the other player from passing is scored as an act of betrayal. Taking a longer route to avoid a gate that may or may not be locked is scored as an act of distrust.

The layout of the microworld is complete. The dialog boxes are being scripted and installed. We expect to be able to discern and differentiate between patterns of behavior that promote trust and cooperation and patterns of recrimination and betrayal. We hope to identify behavior aimed at mitigating the deleterious impact of distrust and betrayal on teamwork and cooperative behavior.

Interim Fourth quarter 2002 progress report

This report discusses work performed at Kansas State University as part of the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. The report has four sections. The first summarizes our activities, findings, and products. The second reviews our experimental paradigm. The third section details the four on-going experiments. The final section is a list of publications and presentations supported and facilitated by the MURI grant.

Summary

Trust is the willingness to allow the uncertain actions of others to determine what happens to you. It is the glue that holds teams (e.g., infantry units) together. The transition to the Objective Force Warrior has the potential to erode the traditional social foundation for the establishment of trust. Erosion of trust can have disastrous consequences for the team and for its performance and effectiveness. The MURI-supported research at Kansas State focuses on identifying and mitigating the impact of the erosion of trust on team performance.

We have completed one experiment and continue to collect data on three additional experiments that use a variety of individual and cooperative cognitive tasks to investigate the impact of the erosion of trust on the performance of two-person teams. We are compiling a literature review on the organizational foundations of trust are preparing a series of journal articles on the experiments and on the foundations and erosion of trust.

Findings

Methodology

We have developed a simple but powerful methodology for investigating trust and teamwork in laboratory settings. Unlike the approaches traditionally used in studies of trust, our methodology does not require the experimenter to induce trust attitudes. Rather,

our participants play an economic game (with real money at stake) which makes them declare their willingness to trust and to be trustworthy. By playing the economic trust game, the two-player team sorts itself into one of three categories of trust – trusting, distrusting, or betraying. This self-selection is a quasi-experimental manipulation of one of the independent variables in our experiments.

Using the economic trust game as an independent variable in a quasi-experiment is a radical departure from the original uses of the trust game. Experimental economists introduced the game to use it as a dependent variable in studies of behavioral deviation from economic theory. Using it as the stimulus in a quasi-experimental design allows us to circumvent the ethical quandaries associated with most experimental studies of trust. It obviates the need for deception or role-playing or the induction of artificial attitudes and affect.

The economic quasi-manipulation of trust works. Teams that sort themselves into different categories perform differently on subsequent teamwork. The full experimental paradigm is presented in a separate section below. The behavioral consequences of the quasi-manipulation are summarized next.

Laboratory studies of trust and teamwork

Teams that sort themselves into the trusting category perform a subsequent cooperative problem solving task more efficiently than teams that sort themselves into the distrusting or betraying categories. This result is in accord with our expectations and demonstrates the viability of our experimental paradigm.

Distrust is more detrimental to team performance than betrayal[1]. We believe this result is the product of the lack of reciprocal feedback that underlies our economic

[1] It is important to note that betrayal in the economic game is the failure to be trustworthy. Betrayal is the self-interested response to an act of trust. The consequences of betrayal in the economic trust game are less extreme than of treachery or traitorous action and should not be construed as such.

implementation of distrust. The argument linking feedback to distrust is presented in the discussion of the experimental paradigm below.

A preliminary review of data from a task with a distinctly unequal distribution of responsibility suggests that a self-evaluation of poor performance in the more demanding job increases the likelihood of a choosing the distrusting outcome on subsequent play of the economic game.

Field experiment

We evaluated the feasibility and utility of conducting an experiment on leader presence and team performance under fire in a paintball range. We observed consistent differences in team performance across the two conditions of leader presence. Teams used less ammunition and responded to commands more quickly when the leader was in the field than when the leader was a voice on the radio. We are taking the steps needed to obtain permission to continue this field experiment at Fort Riley under more realistic conditions. The Fort Riley experiment will add psychophysiologic measures of stress to the behavioral measures of trust and performance.

Products

It is too early to claim to have found a guiding principle that can unequivocally be applied to training or in the field. However, if our findings hold up and are replicated by others, we expect to reveal that full and reciprocal feedback is one of the foundations of trust and, conversely, that the perception of privileged feedback is a foundation of distrust[2]. The guidance this research promises to give the Army would be that maintaining full and reciprocal feedback can mitigate the erosive impact of distrust. We are currently taking our initial steps towards developing and testing a theory of trust and feedback reciprocity.

---

[2] Feedback is reciprocal when both (all) parties have access to all information about each other's actions. In contrast, feedback is privileged when it is perceived to flow only one way.

We expect our work to inform the development of IMPRINT by (1) expanding the taxa of cognitive stressors to include trust, (2) by differentiating between stress degradation algorithms for individuals and teams, and (3) by constructing stress degradation algorithms that can be folded into the existing ACT-R cognitive architecture of IMPRINT.

Experimental paradigm

Overview

Our experimental paradigm is simple yet powerful. Two participants are asked to show up at the same time in different places. They may or may not know each other. In the familiar condition they do; in the anonymous condition they do not. (We expect anonymity to exacerbate the erosive impact of distrust on team performance.) The participants perform four computer-mediated tasks. The first and last are the same cognitively demanding task. The replication affords a repeated measure analysis of individual performance at a task that does not involve teamwork. We expect to see learning effects due to repeated play.

The second task is an economic game, our quasi-manipulation of trust, described below. The game is played for cash which the participants receive immediately. The third task requires some degree of cooperation. In the familiar condition, they are brought into one room to work together on the task using separate computers. In the anonymous condition, they remain in separate rooms. A shorthand characterization for the basic experimental design is AXBA, where A is the repeated task, X is the economic quasi-manipulation of trust, and B is the team task.

This design makes it possible to differentiate the effects of the quasi-manipulation of trust X on the team task B and on the individual task A. If performance on task B is poor (when compared to other teams in other trust conditions) but performance on task A

is not degraded, then we have documented a dissociation in the impact of the quasi-manipulation of trust on team and individual performance. Details follow.

The individual cognitive task

In separate rooms, each participant is asked to perform a task that tests a fundamental cognitive ability. This provides a baseline for the repeated measure of individual cognitive performance. All experiments conducted in 2002 used the mental rotation task, a classic measure of visuospatial ability. Example stimuli are shown in Figure 1. The subject's task is to determine as quickly as possible whether the two figures are the same, differing only by an angle of rotation, or mirror-images.

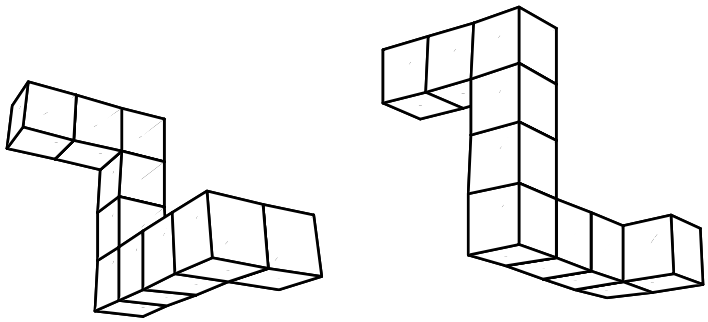


Figure 1. Sample stimuli used in the mental rotation task

The economic trust game

While still in separate rooms, the participants the use the computer network to play the economic trust game, shown in Figure 2. They know that they will play the game once and only once and that real money is at stake. The game has three possible outcomes with different payouts. The choices the players make determine the amount of cash they receive. We treat the outcome of the economic game as an independent variable. This quasi-manipulation enables participants to self-select into one of three categories of trust – distrusting, trusting, or betraying. Participants receive the amounts of cash indicated in Figure 2 immediately upon completion of the game.

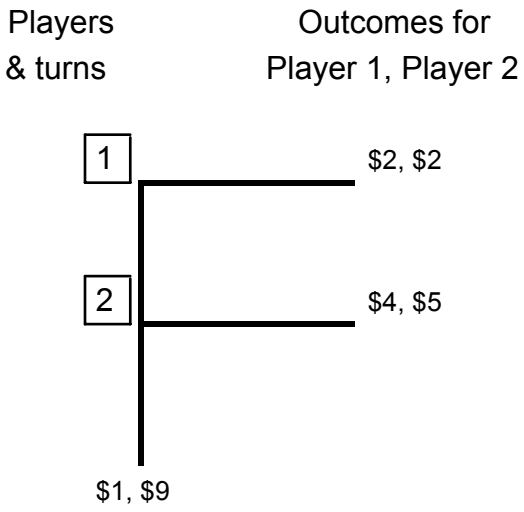


Figure 2. The extensive-form trust game used as the stimulus in the quasi-manipulation of trust (modified from McCabe, Rassenti, & Smith, 1995). The numbers in the squares represent Player 1 and Player 2. Player 1 has the first move. A move to the right is a distrusting move that nets both players $2. A move down is a trusting move that passes play to Player 2. A move right by Player 2 is a trustworthy move that nets Player 1 $4 and Player 2 $5. A move down by Player 2 betrays Player 1's trust.

Player 1 has the first move and two options, to trust Player 2 by moving down and passing play to Player 2, or to distrust Player 2 by moving right[3]. Moving down is an explicit declaration of trust: Player 1 is willingly entering an uncertainty situation in which the actions of another person, Player 2, determine what will happen to them both. It sends a strong signal to Player 2, "If you move right, we will both be better off. Reward me for trusting you." If Player 1 moves right, it is an explicit declaration of distrust. The signal Player 1 sends Player 2 is "I think you are more motivated by your individual self-interest than in our common good."

[3] The distrusting move is the move predicted by economic theory. Theory suggests that Player 1 should infer that Player 2 is likely to be motivated by self-interest and, given the chance to play, will move down to take the $9 rather than the $5 offered by a move to the right. Theory continues by suggesting that Player 1 should use backward induction to infer that giving Player 2 a chance to play will net Player 1 only $1. This outcome is less desirable to Player 1 than the $2 offered moving right. Hence, Player 1 should move to the right.

When Player 1 moves to the right, the game is immediately over. Both Players receive $2. For the purposes of data analysis, we say that the team has self-selected itself into the distrusting category.

If Player 1 moves down, play is passed to Player 2. Player 2 has two options. If Player two right is an explicit act of trustworthiness that rewards Player 1's trust. The game is immediately over. Player 1 receives $4 and Player 2 receives $5. For the purposes of data analysis, we say that the team has self-selected itself into the trusting category. Moving down is an explicit act of betrayal that also ends the game. Player 1 receives $1 and Player 2 receives $9. For the purposes of data analysis, we say that the team has self-selected itself into the betraying category.

Both the trusting and betraying outcomes are associated with full and reciprocal feedback about each other's actions. Both players know the move the other player made. Distrust is the only outcome for which feedback is not reciprocal. A distrusting move makes it obvious to both players the move that Player 1 made, but does not give Player 2 the opportunity to reveal the move that never got made. Only Player 2 knows the move that would have been made, given the opportunity. Player 2 has privileged information. By distrusting, Player 1 sets up a situation with incomplete feedback. We suspect that the lack of feedback reciprocity may underlie the differences in performance in the subsequent task that requires teamwork.

The task that requires teamwork

The third task our participants perform is one two tasks that require teamwork and cooperation. The first is the Hobbits and Orcs problem solving task in which players take turns. This task allocates the responsibility for success (and the opportunities for error) evenly across both players. The second task is 'the Boat Game,' a distributed command and control task in which one of the players has considerably more responsibility (and more opportunities to make an error) than the other. The disproportionate allocation of responsibility allows us to test hypotheses about distrust and feedback reciprocity.

The fourth task for our participants is the replication of first to test the effects of the economic trust game and team performance on individual performance. At the end of the fourth task, participants are debriefed and thanked for their time.

Developments

During the fourth quarter of 2002, we conducted three laboratory experiments and a field experiment, completed the software development for a fourth experiment, and compiled a literature review on the organizational foundations of trust. The first of these experiments is reviewed here. The literature review will be completed and delivered as an article for publication during the first quarter of 2003.

On the effect of play in an economic trust game on subsequent team problem solving (Hobbits and Orcs)

This laboratory experiment applies the methodology discussed above to investigate the effect of the quasi-manipulation of trust on team problem solving and individual cognitive performance. The individual task is mental rotation (Metzler & Shepard, 1971), a classic measure of visuospatial ability. The team problem solving task is a variant of the Hobbits and Orcs river-crossing problem (Greeno, 1974), a classic game from the literature on individual problem solving.



Figure 2. The initial state of the distributed computer simulation of the Hobbits and Orcs game.

Figure 2 is a screen-shot of the initial state of the game. Three Hobbits and three Orcs are on one side of a river and want to cross in a small boat. The goal of game is to move them all safely across the river. Three rules constrain play: 1) the Orcs may not outnumber the Hobbits anywhere at any time (or else the Orcs eat the Hobbits, not a good thing), 2) someone must be in the boat to get it across the river, and 3) no more than two characters may be in the boat at one time. In our variant of the game, one player is responsible for launching the boat from the left bank, the other for launching the boat from the right bank. Dependent measures include the number of moves, the number of illegal moves (e.g., allowing a Hobbit to get eaten), and the amount of time taken. Players are given 100 moves or 45 minutes to complete the game.

The four treatment conditions fully cross team-member presence at two levels (face-to-face, anonymous) and communication at two levels (free talk, no talk). We view distrust as a source of stress that is likely to have a deleterious impact on team problem solving. We expect both anonymity and the inability to communicate to exacerbate this impact. Table 1 enumerates the distribution of teams in each of the four conditions. Data collection is nearly complete.

Table 1. The distribution of teams in four treatment conditions in the team problem solving task. D - distrusting, T – trusting, B - betraying

		Presence					
		Familiar			**Anonymous**		
Communication	**Free talk**	Same room, ordinary conversation			Different rooms, radio communication		
		D	**T**	**B**	**D**	**T**	**B**
		2	5	4	3	4	1
	No talk	Same room, no conversation allowed			Different rooms, no communication		
		D	**T**	**B**	**D**	**T**	**B**
		3	4	3	5	8	6

Figure 3 shows a sample of the data from the Hobbits and Orcs experiment. The experimental condition is familiar-free talk (upper right quadrant of Table 1); the dependent measure is total time to complete the task. Teams that self-selected into the trust category completed the task faster than teams that experienced either betrayal or distrust. Distrusting teams fared least well.

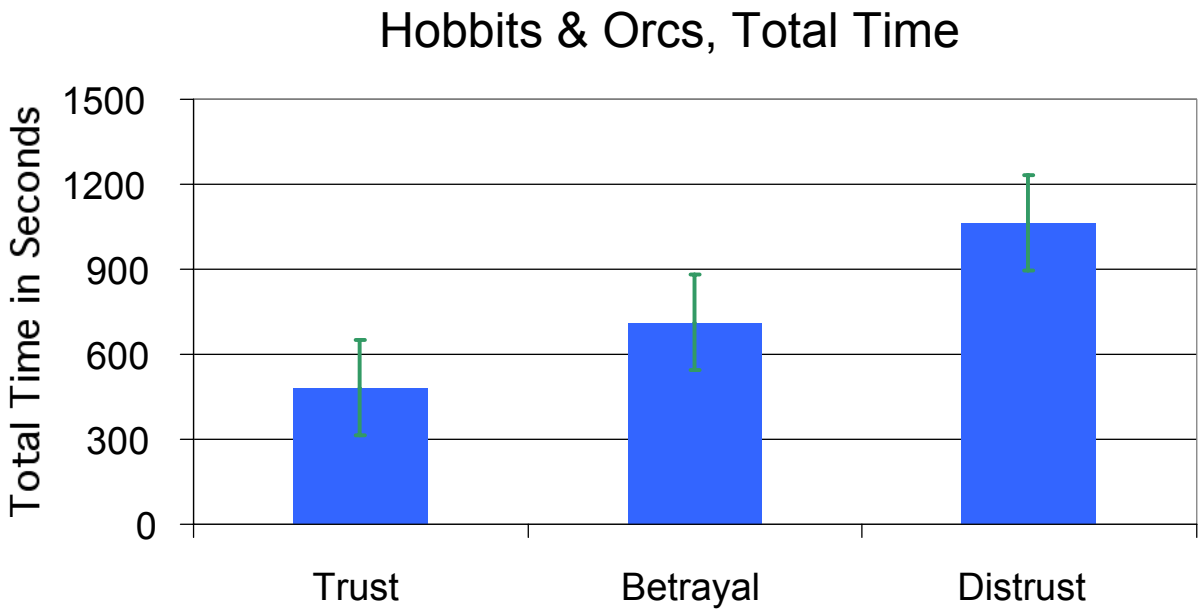


Figure 3. Sample time-on-task data from the familiar-free talk condition of the Hobbits and Orcs experiment.

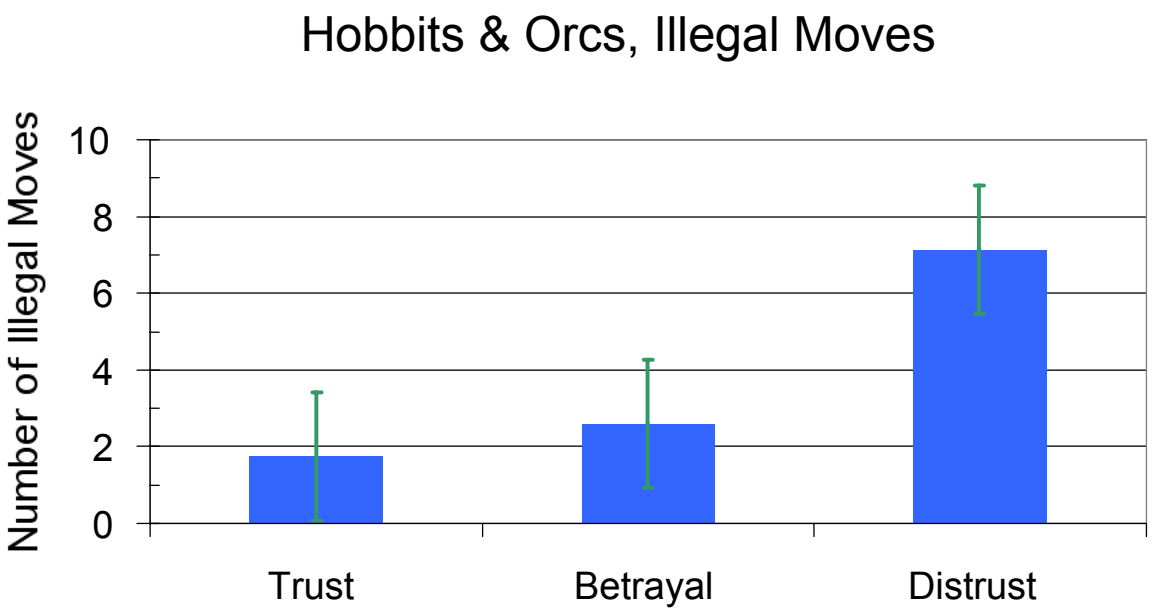


Figure 4. Sample data from the familiar-free talk condition of the Hobbits and Orcs experiment, number of illegal moves.

Figure 4 shows the number of illegal moves made by teams in the familiar-free talk condition. Once again, teams that self-selected into the trust category fared best and distrusting teams fared least well.

On the mental-rotation task, most participants benefit from repeated play, getting both faster and more accurate on the second trial. This improvement appears to be impervious to the outcomes of the economic trust game and the problem solving game. Players from trusting pairs perform no better than players from distrusting and betraying pairs.

In sum, the individual cognitive processes involved in mental rotation of spatial objects are unaffected by our laboratory quasi-manipulation of trust. In contrast, team performance is strongly affected by the quasi-manipulation. Teams that trusted each other before working together proceeded to work together best. Teams that experienced distrust proceeded to find it difficult to work together. Teams that experience betrayal occupy a middle ground. This combination of results indicates there is a strong dissociation in the effects of levels of trust on individual cognitive performance and

teamwork. The stress degradation algorithms for the effects of trust will have to differentiate between effects on individuals and effects on teams.

On the effect of play in an economic trust game on subsequent team performance in a distributed, cooperative task with differential responsibility (The Boat Game)

This laboratory experiment investigates the generality of the findings from the Hobbits and Orcs experiment by using a different team task, called the Boat Game. As in the Hobbits and Orcs experiment, the four treatment conditions fully cross presence at two levels (face-to-face, anonymous) and communication at two levels (free talk, no talk). The economic trust game, Figure 2, is again used as the stimulus in a quasi-manipulation of trust.

Figure 5 is a screen shot of the stimuli in the Boat Game. The green represents land, the blue a river, and the red a boat moving upriver. The goal is to get the boat up the river to a dock as quickly as possible. What makes the game difficult is that the two players share control of the boat. One player is the captain responsible for steering the boat. The other player is the engineer responsible for its velocity. The coordination is made complex by the significant time lag between the issuing of an instruction (to turn, to speed up, to slow down) and its realization.

The game has a monetary incentive. At the end of the game, when the players get to the finish line, they split the total of the dollar amounts shown on the “fuel gauge” and the “clock.” The clock counts down from $15.00 at the rate of $0.01 per second; in Figure 5 it shows $12.26. The fuel gauge starts at $15.00 and counts down by $0.05 with each instruction to increase speed; in Figure 5, it shows $11.15. If the engineer boosts the velocity too much, it is simply impossible for the captain to turn the boat sharply enough to avoid running aground. Running aground is expensive. The boat comes to a full halt for five seconds and then pops back out to the center of the river at full stop.

The pilot has the easier job. Steering the boat is relatively simple when the boat is not moving too fast. The engineer shoulders all the responsibility both for velocity and for running out of fuel. It is easy to run out of fuel. Running aground requires burning some

fuel to get underway again. Run aground too often, and there will soon be little fuel left. The trick to successful play is to burn only enough fuel to reach the fastest velocity the pilot can handle and then burn no more.

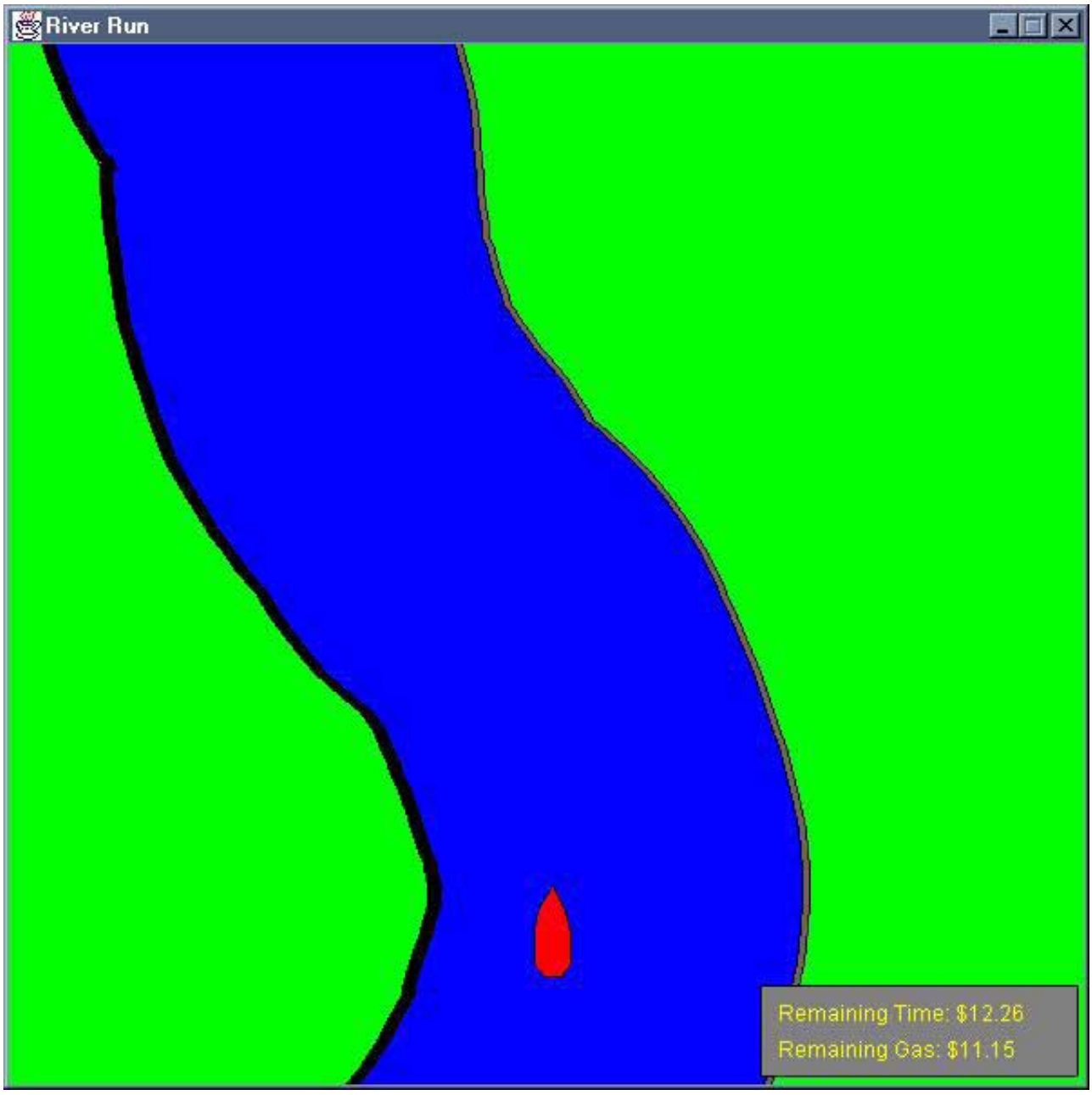


Figure 5. The computer screen in the distributed, cooperative task with differential responsibility, a.k.a. the Boat Game.

The differential responsibility of the Boat Game stands in contrast to the exchange of turns in the Hobbits and Orcs game. At issue in this experiment is the impact of differential responsibility on the quality of cooperation. Inferential data analysis is underway and will reported in the first quarter of 2003.

On the effect of team performance in a distributed, cooperative task with differential responsibility on subsequent play in an economic trust game

This laboratory experiment changes the sequences of tasks outlined in the section on Experimental Paradigm. In this experiment, the Boat Game is played before the economic trust game. At issue in this experiment is the impact of the quality of cooperation (money received) during play of the Boat Game on subsequent play of the economic trust game. Inferential data analysis is underway and will reported in the first quarter of 2003.

Paintball

We came, we saw, we conquered. During the fourth quarter of 2002, we demonstrated that it is possible to conduct an experiment on teamwork and leader presence under live paintball fire and get replicable results. The paintball court was set up like a firing line with alternating bunkers on each side and an enemy strongpoint at the far end. One experimenter acted as the team leader; another was the 'enemy' behind the strongpoint. Two participants were 'soldiers' armed with paintball guns and 50 rounds of ammunition. One was on the right side of line, the other on the left. The 'enemy' had a gun and unlimited firepower. The soldiers' task was to follow the leaders' orders either to advance or to cover the advancing teammate. While advancing, a soldier had to capture flags placed in the center of the lane. Capturing a flag required trust in both the teammate, to provide protective fire, and in the leader, to indicate when it was safe to advance.

The experimental manipulation was leader presence. In one condition the leader was in the firing line with the soldiers. In the other, the leader was a disembodied voice on a radio. This manipulation was designed to capture one of the salient differences between

the function of infantry units in the traditional Army and the Army of the Objective Force Warrior. Dependent measures were the amount of ammunition used, the time to capture flags, and subjective reports of stress and anxiety.

The two soldiers used less ammunition and responded to the leader's instructions more quickly when the leader was in the field of fire. This result has led us to move through ARO and ARL channels to secure permission to continue the experiment indoors at Fort Riley. Fort Riley is adjacent to Manhattan Kansas, the home of Kansas State University. We expect to obtain permission by January 2003 and to commence data collection in February, 2002.

Publications and presentations

MURI-Supported Work

Thesis

Lauren L. Murphy. (2002). Physiological and behavioral measures of workload: An empirical investigation of operational errors in a simulated air traffic control environment. Unpublished M. S. thesis, Kansas State University.

In preparation

Murphy, L., & Smith, K. A laboratory study of task demand, response to task demand, and operational errors. Journal of Experimental Psychology, Applied.

Smith, K., Shane, A., Kearney, A., & Swan, B. The impact of an economic quasi-manipulation of trust on team problem solving. Science.

Proposed submissions to a special MURI issue of Military Psychology

Smith, K., Sitzes, R, & Freund, S. The foundations and erosion of trust in teams.

Smith, K., Bolton, J., Kearney, A., Pruitt, N., Shane, A., Swan, B., Toll, M., & Younger, E. Hobbits and Orcs and Xboat game.

Smith, K., Hawkins, S., & Bloomberg, S. The unequal distribution of responsibility is a behavioral basis for distrust.

Pangburn, K., Smith, K., & Pangburn, H. Leader presence on team performance under fire on a paintball lane.

Peer-reviewed presentations

Swan, B., Shane, A., Ratcliff, E., Pruitt, N., Kearney, A., Freund, S., Fosse, N., & Smith K. (2002). The impact of an economic manipulation of trust on team problem solving. The 2002 Annual Meeting of the Society for Judgment and Decision Making. Kansas City, MO. Nov 24-25.

Fosse, N. & Smith, K. (2002). The effect of trust on cooperative behavior. The 2002 Sigma Xi Student Research Conference. Galveston, Tx. Nov 15-17.

Fosse, N., Murphy, L., Perry, J., Ratcliff, E., & Smith, K. (2002). Preliminary studies of the impact of an economic manipulation of trust on team problem solving. Oklahoma-Kansas Society for Judgment and Decision Making. Oklahoma City, OK. April 27.

MURI-Facilitated Work

Publications

Smith, K., Dickhaut, J., McCabe, K., & Pardo, J. (2002). Neuronal substrates for choice under ambiguity, risk, gains, & losses. Management Science **48** (6), 711-718.

In review

Dickhaut, J., McCabe, K., Nagode, J. C., Smith, K., & Pardo, J. (in review) The role of context in choice: The effect of different comparison gambles on risky choice. Proceedings of the National Academy of Science.

Peer-reviewed presentations

Smith, K. & Dickhaut, K. (2002). Auction structure, heart rate, and neuroeconomic behavior. Minnesota Conference on Neuroeconomics. Minneapolis, MN. Oct 25-27.

Smith, K., Dickhaut, J. & Peck, N. (2002). Affective determinants of economic behavior. The 9th Behavioral Decision Research in Management Conference. Chicago, IL. May 31-June 1.

In preparation

Rustichini, A., Dickhaut, J., Ghirardato, P., Smith, K., & Pardo, J. A brain imaging study of procedural choice. Econometrica.

Smith., K., & Dickhaut, J. Auction structure, heart rate, and neuroeconomics. Journal of Experimental Economics.

Smith., Lewin, J., and Hancock, P. A. The risk space for collision detection. Journal of Experimental Psychology, General.

Smith, K., Scallen, S., Knecht, W. R., Hancock, P. A., & Avans, D. Determinants of self-separation in free flight. Human Factors.

Progress Report
May 22, 2002

Florida Institute of Technology
Prof. John Deaton

HOLMES REGIONAL MEDICAL CENTER
FIRST FLIGHT AIR AMBULANCE EVALUATION
START DATE: FEBRUARY 18, 2002
(TIME WORKED: 180 HOURS AS OF MAY 18)

Background Research

The first month of the project was spent doing background research at the libraries of Florida Tech Melbourne Campus, Brevard Community College Palm Bay Campus, and University of Central Florida Orlando Campus. The purpose of this research was to develop guidelines for performing a human factors evaluation of any given program, which would be used as a template for the upcoming First Flight program evaluation. This research also focused on procuring information on air ambulance programs throughout the United States and the world to determine whether a similar evaluation had ever been done at any other air ambulance operation. The information that was discovered indicated that no such evaluation had been done as a stand-alone document. However, later research turned up the possibility that all of the relevant aspects of a human factors evaluation could likely be found in the documentation that most programs had to complete for CAMTS certification.

This phase of research turned up a wealth of information on human factors evaluations, and allowed for the formulation of a definitive document on evaluation guidelines that would be used as a template for the evaluation to follow at First Flight.

Heath First Research

The first day at First Flight was March 18. That week was spent mainly getting to know the entire First Flight staff, a step that would be critical. The objectives of the evaluation would require that the staff be comfortable with my presence and with me observing them in their everyday duties and interactions. It was determined that it would be imperative that I spend as much time as was necessary to accomplish this, making the staff aware that the proposed evaluation was not in any way threatening to them professionally.

Once I was established as a part of First Flight where the staff felt comfortable with me being there, plans were laid out regarding when I would be there for the most part. I retained a flexible schedule for purposes of observing staff interaction at different times and under different circumstances, but I also established a schedule of times that I would be there. I also took that time to make the staff aware of what information I was interested in gathering, and how it would be used in the evaluation. This time involved constant reassurance of the staff that the program evaluation was not threatening to them professionally or personally, and that all information gathered for the evaluation would remain confidential.

Progress Report
May 22, 2002

Florida Institute of Technology
Prof. John Deaton

Informal Interviews

The initial stage of my information gathering involved talking at length to some of the staff to obtain a general idea of the history of First Flight. History information was relatively easy to come by, being documented in program brochures and newspaper articles. Moreover, I was able to informally interview some staff members, and in particular Mrs. J Willette (Chief Flight Nurse/ Asst. Program Director), and Mr. M. Briganti, a flight paramedic who has the distinction of being the longest serving First Flight team member, having been in the program since its inception in 1988. Mrs. Willette has only been with the program for almost 5 years, but in that time has been largely responsible for helping to give the program the positive direction and influence that it presently enjoys.

History

Since the program's inception, it has grown significantly and consistently in the number of patient transports per year, as well as in staff size and organizational complexity. Of greatest pride to Ms. Willette is that the program's reputation and public perception has increased dramatically in the time that she has been here. She came to First Flight at a time when the program's image within the medical community was under some pressure because of what was considered to be grossly unprofessional behavior by the crewmembers. Since then the program standards and image have changed for the better. This was substantiated by Mr. Briganti, who indicated that the present standards of entry criteria, experience, training and continuing education have all seen continuous improvements over the life of the program, but most dramatically within the last few years, under Mrs. Willette's watch. Mr. Briganti expressed that the staff are now more prepared to deal with the highly charged situations that can arise in the completion of their duties in a professional manner. They have a more efficient support structure (management and organizational leadership) that makes it possible for them to be more professional in their dealings with the public, other organizations, and even other departments within Health First.

Mr. Briganti also, as the longest serving member of the First Flight team, was qualified to document the considerable improvement in industry standards and equipment availability. He spoke of the influence that rising industry standards have had in making First Flight and all other air ambulance operations more effective in patient care. First Flight has been positioned at the leading edge of this technological advance, with every effort made to have available to the crews the latest or best possible equipment, and management support of crew training to maintain currency with industry trends.

During this time I also watched a new Flight Paramedic undergoing training on the day shift. His training was highly task oriented, and he learned by doing all the tasks that

would be required of him as an active team member, under the supervision of one of the senior flight paramedics. The training was very focused, and the new crewmember was specifically educated on the manner of completing documentation, including appropriate terminology to be used as a part of each document.

Existing Documents

I spent approximately four weeks going over First Flight documents, as well as interacting more and more with staff. I started off with lessons in the basics of flight physiology, and a ground safety helicopter course so the in the event that there was an opportunity to do so, I could be allowed to participate in a flight under the ride-along program. This would give me a hands-on experience of crew interactions and communications in flight. The ride-along program is an idea that is gaining popularity among air ambulance programs as a public awareness tool. Although participants are usually limited to personnel that are in some way affiliated with Emergency Management Services, on occasion other personnel are allowed to ride as a part of the program. The First Flight public relations program also sees the crews go into the community to schools and other civic organizations to make presentations and promote public awareness on various health issues and safety programs.

Examining the First Flight document cabinet also provided another source of information into the history of the program, as well as allowing access to information on personnel requirements and orientation, program policies, procedures and protocols, aeromedical training issues and procedures, safety procedures, inservice requirements and continuing education units, and public relations and special events. The staff meeting minutes provided considerable information on the issues that led to action items within First Flight, and procedures for accomplishing those action items.

Future Plans

The next step will involve administering surveys to the First Flight staff to obtain a formal idea of crew and pilot perceptions of job/task/workspace conditions, design, and demands. This is expected to provide an idea of what aspect of human factors concerns should be looked at first, and most intensively, in the evaluation. The question bank for these surveys has come from numerous sources, including the examination of meeting action items, and conversations with staff regarding job and task design and demands. Some of the questions will be identical to those asked on an informal basis, and will provide a formalization of the information gathered in those conversations for purposes of analysis.

Quarterly Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield

Army MURI, Subcontract (*Models of Workload Transitio*) to U. of Central Florida

Gerald Matthews and Joel S. Warm

University of Cincinnati

Performance Period: 9/01/02-12/31/02

During this period, two studies were accepted for publication, data analyses and write ups were completed for four studies to be presented to the Annual Conference of the Human Factors and Ergonomics Society, Baltimore MD and Drs. Matthews and Warm presented invited papers at a research conference sponsored by the US Army Research Laboratory, US Army Research and Engineering Division, Aberdeen Proving Ground, MD.

Publications

1. *James L. Szalma, Joel S. Warm, Gerald Matthews, William N. Dember, Ernest M. Weiler, Ashley Meier, and F. Thomas Eggemeier*, Effects of Sensory Modality and Task duration on Performance Workload, and Stress in sustained Attention. **Human Factors** (in press).

The expanded use of automation has altered the role of operators in many work settings from active controllers to executive-monitors who must attend to displays and take action only in the event of imminent problems. There is substantial evidence to indicate that monitoring tasks impose high workloads on operators and that they find such tasks to be stressful. Overall workload scores in vigilance typically fall in the mid-to upper-range of the NASA-TLX, a standard scale for measuring perceived mental workload, and increase linearly over time (Warm, Dember, & Hancock, 1996). Moreover, observers report being less attentive and more bored, strained, irritated, and fatigued at the end of a vigil than prior to its start (Warm, 1993). To date, research on the workload of vigilance has been restricted to visual tasks and stress research has not examined the rate of gain over stress over time. The sensory modality of signals and time on task are key psychophysical considerations in vigilance research. Accordingly, this study was designed to examine the effects of the sensory modality of signals and time on task on the workload and stress of sustained attention. Observers monitored the repetitive presentation of white bars on a VDT or bursts of white noise presented binaurally via headphones for occasional changes in stimulus duration. Pilot work insured that the auditory and visual discriminations were equated for difficulty under alerted conditions. Workload and stress were measured after 10, 20, 30, or 40 min on watch in a between groups design. Workload was assessed by means of the NASA-TLX, while the Dundee Stress State Questionnaire (DSSQ), a multidimensional instrument for assessing transient states associated with mood, arousal, and fatigue (Matthews et al., 1999; 2002) indexed stress. Signal detections declined over time and were higher for auditory than for visual signals. Overall workload scores on the NASA-TLX were similarly substantial for both sensory modalities with scores on the Frustration scale increasing linearly over time. Results with the DSSQ indicated that the stress of sustained attention is multidimensional in character and that it is tied to both the sensory modality of signals and time on watch. Observers in the visual task were found to be uniformly less motivated and happy than those in the auditory task. In addition, the levels of tension and task-relevant and task-irrelevant cognitive interference increased monotonically over the course of the vigil for visual signals but followed and inverted U-shaped function for auditory signals, rising during the early portion of the vigil and returning to baseline by the end of the session. In order to minimize operator stress, it is evident that display modality and time on task should be important considerations in the design of systems that include vigilance components.

2. *Rebecca A. Grier, Joel S. Warm, William N. Dember, Gerald Matthews, Traci L. Galinsky, James, L., Szalma, and Raja Parasuraman*, The Vigilance Decrement Reflects Limitations in Effortful Attention, Not Mindlessness. **Human Factors (**in press).

Robertson et al. (1997) have proposed that detection failures in vigilance tasks result from a "mindless" withdrawal of attentional effort from the monitoring assignment. To explore that view, they

modified the traditional vigilance task in which observers make button-press responses to signify the detection of rarely occurring critical signals to one in which button-press responses acknowledge frequently occurring non-signal events and response-withholding signifies signal detection. This modification is designed to promote a mindless withdrawal of attentional effort from the task through routinization. The current study challenges the "mindlessness" position. Observers monitored the repetitive presentation of a display consisting of a large centering dot with two smaller dots normally located at equal distances to its right and left. Critical signals for detection were cases in which the distance of one of the flanking dots was further from the centering dot than the other. Observers performing the task in the normal manner rated it high in workload (NASA-TLX) and stress (Dundee Stress State Questionnaire) and this effect was duplicated with Robertson et al. modification. In addition, observers in both the standard and modified formats revealed concern over performance efficiency as reflected in the Task Relevant Interference scale of the DSSQ, a result that is inconsistent with the view that observers withdraw attention from the vigilance task. Contrary to the mindlessness model proposed by Robertson and his associates, the present results support a resource view of attention, wherein vigilance tasks are construed as attentionally demanding, thereby exhausting information-processing resources (Matthews, Davies, Westerman, & Stammers, 2000; Warm, Dember, & Hancock, 1998; Warm & Dember, 1998). Efforts to enhance monitoring performance need to find ways to reduce the information-processing load on operators; strategies designed to guide operators in maintaining proper attentional focus are not likely to be the solution.

Paper Presentations

1. *Carolyn R. Alikonis, Joel S. Warm, Gerald Matthews, William N. Dember, and James Kellaris*, Effects of Music on the Workload and Stress of Sustained Attention. Paper presented at the HFES, Baltimore MD, September 30-October 4, 2002.

Two models of the high workload associated with vigilance tasks are the direct-cost and indirect-US Army Research Laboratory, US Army Research and Engineering Division, Aberdeen Proving Ground, MD.cost views. The former attributes high workload to the need for continuous observation in discriminating signals from neutral events; the latter attributes it to efforts to combat the boredom associated with monotonous vigilance tasks. A recent study from our laboratory (Hitchcock, et al., 1999) tested these opposing views using a simulated air-traffic control display in which observers monitored planes passing over a city for collision headings during a 40-min vigil. Observers were presented with reliable cueing which rendered observation necessary only when low probability critical signals were imminent. In this way, cueing lowered the need for continuous observation of the display but did not eliminate the tedium of vigilance, observers remained in a monotonous environment with little to do. Consequently, the direct-cost model led to the anticipation that cueing would result in a high-boredom low workload profile while the indirect-cost model led to the prediction that cueing would result in a high-boredom, high-workload profile. The results supported the direct-cost view that the workload of vigilance is task-induced. In comparison to a non-cued control group, cueing lowered the perceived workload of the vigil on the NASA-TLX but had no effect upon boredom as indexed by the Task-Related Boredom Scale (TBS:Scerbo, 1998). The present study extended this line of investigation by disengaging workload and boredom in still another way. Monitors were permitted to listen to a musical selection, *Heart Zones*, designed to reduce boredom and stress while they monitored the simulated air-traffic control display. Although the musical selection targeted boredom, it did not lessen the need for careful observation of the vigilance display, observers still had to make continuous signal-noise discriminations during a 40-min vigil. Consistent with the direct-cost view, the musical selection lowered rated boredom on the TBS but had no effect upon perceived workload on the NASA-TLX. These studies indicate that efforts to combat high workload in automated systems requiring substantial monitoring by operators should focus upon task-related determinants.

2. *William S. Helton, Joel S. Warm, Gerald Matthews, Kevin J. Corcoran, & William N. Dember,* Further tests of an abbreviated vigilance task: Effects of signal Salience and Jet Aircraft Noise on Performance Efficiency and Stress. Paper presented at the HFES, Baltimore MD, September 30-October 4, 2002.

Because of their demanding character, vigilance tasks have been used in test batteries to assess the effects of stressors such as infectious disease and drugs on performance efficiency. In addition, vigilance

tasks have been employed as tools to examine the impact of a variety of clinical entities on attention, including brain damage and emotional disorders. Although useful in these roles, the long duration of traditional vigilance tasks has often been an inconvenience. Accordingly, a recent study from our laboratory (Temple, et al., 2000) developed an abbreviated 12-min computerized vigilance task in which long-duration tasks are precluded (e.g., performance assessment batteries, neuropsychological testing, and brain imaging. The task requires observers to monitor the repetitive presentation of light gray capitol letters consisting of a D, O, and a backwards D in which the letter O is the critical signal for detection. The letters are presented at a rate of 57.5 events/min with a dwell time of 40 ms and are exposed against a visual mask consisting of unfilled circles on a white background. The initial work with this task demonstrated that it reproduces the vigilance decrement, high workload (NASA-TLX), and stressful character (DSSQ) of monitoring tasks lasting 30min or more. In addition, the abbreviated task was found to duplicate the signal salience effect, a major finding associated with long-duration vigilance tasks and performance efficiency on the abbreviated task was enhanced by caffeine – a drug that benefits long-duration tasks. The present study was designed to explore further the similarity of the abbreviated task to long-duration vigils by investigating the effects of aircraft engine noise (95dB A at the observer's ear) on performance and stress. As is the case with longer vigils, signal detection in the abbreviated task was poorer for low salience than for high salience signals and stress scores, as indexed by the DSSQ, were generally greater when observers were required to detect low as compared to high salience signals. Unlike longer vigils, however, signal detection in the abbreviated task was superior in the presence of noise than in quiet and noise generally attenuated self-reports of stress. The beneficial effect of jet-aircraft noise for the abbreviated task differentiates it from longer vigilance tasks and supports Poulton's (1977) suggestion that noise may have short-term positive value in vigilance.

3. *Todd D. Hollander, Joel S. Warm, Gerald Matthews, William N. Dember, Edward M. Hitchcock, Christina A. Beam, Lloyd D. Tripp, & Raja Parasuraman*, Cerebral Hemovelocity in Vigilance: Effects of Signal Salience and Periodicity. Paper presented at the HFES, Baltimore MD, September 30-October 4, 2002.

The signal regularity effect – enhanced performance efficiency when critical signals for detection appear in temporally regular as opposed to a temporally irregular manner has a long history in vigilance research. Nevertheless, the precise psychophysical conditions under which this effect can be elicited have not been identified. Toward that end, this study made use of our simulated air traffic control display under conditions in which the contrast ratio of the planes to their background was high (high salience) – the planes were easy to see or the contrast ratio of the planes to their background was low (low salience) – the planes were difficult to see. Using signal detection theory metrics (d' & c), the study demonstrates that the signal regularity effect is rooted in sensing rather than decision-making factors and that the effect is limited to signals of low salience, perhaps because the effort needed to generate veridical temporal expectancies is unnecessary with high salience signals. In addition to performance measures, the present study made use of transcranial Doppler Sonography (TCD), a non-invasive neuroimaging technique employing ultrasound, to monitor cerebral blood flow in the left and right middle cerebral arteries while observers monitored the air-traffic control display. The TCD procedure has served as a metabolic index of information processing in several cognitive tasks and has provided useful insights into cerebral organization in those tasks (Stroobant & Vingerhoots, 2000). It has also served in that capacity in recent studies of vigilance in our laboratory which have demonstrated that several vigilance effects are lateralized to the right cerebral hemisphere (Hitchcock, et al., 2003; Mayleben, 1998). Consistent with those findings, the results of this study indicate that the signal regularity effect is also lateralized to the right cerebral hemisphere.

4. *Christina Beam, Joel S. Warm, Gerald Matthews, Edward M. Hitchcock, William N. Dember, Lloyd D. Tripp, & Raja Parasuraman,* Effects of Sensory and Cognitive Vigilance Tasks on Cerebral Blood Flow Velocity and Subjective Stress States. Paper presented at the HFES, Baltimore MD, September 30-October 4, 2002.

Previous studies from our laboratory using the TCD procedure to provide a metabolic index of information processing during a vigilance task (Hitchcock, et al., 2003; Hollander, et al., 2002; Mayleben, 1999) have all required observers to make sensory discriminations to detect critical signals. However, vigilance tasks can also require discriminations that are more symbolic than sensory in character. Such

tasks are termed cognitive vigilance tasks and nothing is known about blood flow changes with these types of tasks. Accordingly, this study measured TCD-determined blood flow during the performance of two cognitive tasks (simple and complex), as well as a sensory task. All three tasks required observers to monitor a display consisting of digit pairs in which the digits were equal in physical size during a 40-min vigil. In the simple cognitive task, critical signals for detection were cases in which the digits differed by no more than ± 1. In the complex task, critical signals were cases in which the digits met the difference rule and also summed to a value between 4 and 14. In the sensory task, pairs in which the physical size of the digits was unequal constituted the critical signals for detection. Performance was measured in terms of the percentages of signals correctly detected and response time to correct detections. Correct detections declined significantly over time in the sensory condition but remained stable over time in both cognitive conditions. Response times increased significantly over the course of the watch in both the sensory and simple cognitive conditions but were temporally stable in the complex cognitive condition. Blood flow in the right middle cerebral artery paralleled performance as indexed in terms of the accuracy of signal detections for all three tasks but not the speed of detections. The results are consistent with prior neuroimaging findings of right hemisphere dominance in vigilance, but not with arguments that signal detections and response time are equivalent measures of the same underlying process (Buck, 1966). In addition to blood flow measurements, this study also examined stress reactions to the cognitive as compared to the sensory vigilance task by means of the DSSQ, since prior studies of the stress of sustained attention have been restricted to sensory vigilance tasks. The results indicated that the cognitive tasks and the sensory tasks elicited the same pattern of stress reactions found in several other studies employing the DSSQ to assess the stress of sustained attention (Matthews et al., 2002). Specifically, Distress increased and Task Engagement decreased pro pre-to-post vigil. While the three tasks were characterized by qualitatively different discriminations, they shared the requirements that observers make continuous signal/noise discriminations under conditions of uncertainty and lack of control. These requirements have been identified as common threads in eliciting the stress of sustained attention (Hancock & Warm, 1987; Warm, Dember, & Hancock, 1996).

Invited Presentations

Investigators at the US Army Research Laboratory, US Army Research and Engineering Division, Aberdeen Proving Ground, MD are concerned with assessing workload and stress in army personnel and in determining when such personnel may need rest and replacement. They are considering the NASA-TLX, DSSQ, and blood flow measures that have been featured in our research and invited us to present at a workshop that they organized. Dr. Matthews discussed his work on the DSSQ and the multidimensional theory of stress and led a discussion on these issues. Dr.Warm, described the workload and blood flow results from the vigilance studies performed in our laboratory.

Ongoing Studies

1. Data collection was initiated on a study examining the effects of abrupt transitions in signal salience on signal detections and stress in the performance of our abbreviated vigilance task. The study will be completed shortly and described in the next quarterly report.

2. Data collection was initiated on a study assessing cerebral hemodynamics in the abbreviated vigil using the TCD procedure and a measure of cerebral oximetry. The study will be completed shortly and described in the next quarterly report.

3. Data collection was continued on a study using the signal regularity effect to extend our earlier work challenging Robertson et al's (1997) "mindlessness" model of vigilance. The study will be completed shortly and described in the next quarterly report.

STANDARD

NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION

AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

AGREEMENT TITLE: Operator Under Stress Research

AGREEMENT NUMBER: NCRADA-**NAVAIR ORLANDO TSD-02-037**

AGREEMENT ADMINISTRATORS:

NAVAIR ORLANDO TSD

Technology Transfer Office:	**Ms. Patricia R. Nimmo, AIR 49T, (407) 380-8225**
Legal Counsel:	**Robert W. Adams, AIR 772P, (407) 380-8211**
Principal Investigator:	**Ron Wolff, AIR 4962, (407) 380-4583**

UCF MURI / OPUS

Preferred Contact:	**Mrs. Kim Smith, UCF Office of Research, (407) 823-3062**
Principal Investigator:	**Dr. Peter Hancock, Director, UCF MURI/OPUS Laboratory, (407) 823-2310**

This Page Intentionally Left Blank

TABLE OF CONTENTS

PREAMBLE

Article 1.
DEFINITIONS 2

1.1 "Agreement" 2
1.2 "Classified Information" 2
1.3 "Collaborator" 2
1.4 "Controlled Unclassified Information (CUI)" 2
1.5 "Cooperative Work" 2
1.6 "Data" 2
1.7 "Effective Date" 2
1.8 "Exclusive License" 2
1.9 "Government" 2
1.10 "Government Purpose Rights" 2
1.11 "Information" 2
1.12 "Intellectual Property" 2
1.13 "Invention" 2
1.14 "Invention Disclosure" 3
1.15 "Made" 3
1.16 "Militarily Critical Technologies (MCT)" 3
1.17 "Non-Subject Data" 3
1.18 "Non-Subject Invention" 3
1.19 "Patent Application" 3
1.20 "Principal Investigator (PI)" 3
1.21 "Proprietary Information" 3
1.22 "Restricted Access Information" 3
1.23 "Subject Data" 3
1.24 "Subject Invention" 3
1.25 "Tangible Property" 4
1.26 "Unlimited Rights" 4

Article 2. OBJECTIVES 4

Article 3. RESPONSIBILITIES 4

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities 4
3.2 **UCF MURI / OPUS Laboratory** Personnel and facilities 5
3.3 Security Regulations and Directives 5

Article 4. REPRESENTATIONS AND WARRANTIES 5

4.1 **NAVAIR ORLANDO TSD** Representations and Warranties 5
4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Representations and Warranties 5
4.3 Joint Representations 6

Article 5. FUNDING 7

5.1 Payment Schedule..... 7
5.2 Insufficient and Excess Funds..... 7
5.3 No New Commitments..... 8
5.4 Accounting Records..... 8

Article 6. REPORTS AND PUBLICATIONS..... 8

6.1 Final Reports..... 8
6.2 Agreement to Confer Prior to Publication or Public Disclosure..... 9
6.3 Classified Information 9

Article 7. INTELLECTUAL PROPERTY..... 9

7.1 Data..... 9
7.1.1 General Provisions Applying to All Data..... 9
7.1.1.1 Ownership..... 9
7.1.1.2 No Implied License..... 9
7.1.1.3 Marking of Data..... 10
7.1.1.3.1 Data Provided With Less Than Unlimited Rights.. 10
7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information..... 10
7.1.1.3.3 Data That are Subject to 35 USC 205..... 11
7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted..... 11
7.1.1.4 Protection of Data..... 11
7.1.1.5 Release of Data Under the Freedom of Information Act..... 11
7.1.2 Subject Data..... 12
7.1.2.1 Delivery of Requested Subject Data..... 12
7.1.2.2 Rights in Subject Data..... 12
7.1.3 Rights in Non-Subject Data..... 12
7.2 Copyrights..... 13
7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory**..... 13
7.2.2 Copyright License to the Government..... 13
7.2.3 Copyright Statement..... 13
7.3 Trademarks and Service Marks..... 13
7.3.1 Ownership of Trademarks and Service Marks..... 13
7.3.2 Obligation of Employees to Report Trademarks and Service Marks.... 13
7.3.3 Obligation of Collaborators to Notify Each Other..... 13
7.3.4 Responsibility for Filing an Application for Trademark or Service Mark..... 14
7.3.5 License to Use Trademark or Service Mark..... 14
7.4 Subject Inventions..... 14
7.4.1 Obligation to Report Subject Inventions..... 14
7.4.1.1 Collaborators' Instruction to Employees..... 14
7.4.1.2 Timely Invention Disclosure by Inventors..... 14
7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator..... 14
7.4.2 Determination of Subject Inventions..... 14
7.4.3 Title to and Ownership of Subject Inventions..... 14

7.4.4 Filing of Patent Applications..... 15
7.4.4.1 Filing of Patent Applications on Solely Made Inventions..... 15
7.4.4.2 Filing of Patent Applications on Jointly Made Inventions..... 15
7.4.4.3 Preserving Intellectual Property Rights..... 15
7.4.4.4 Filing Deadlines..... 15
7.4.4.5 Copies and Inspection..... 15
7.4.4.5.1 Copies of Prosecution Papers..... 15
7.4.4.5.2 Access to Patent Application File and Right to Make Copies..... 16
7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application..... 16
7.4.5 Nonexclusive License to Subject Inventions..... 16
7.4.5.1 Nonexclusive License Grant..... 16
7.4.5.2 Confirmatory Nonexclusive License Agreement..... 16
7.4.6 Option for Exclusive License to Subject Inventions..... 16
7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement. 17
7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions..... 17
7.4.8.1 Exclusive Licenses and Exclusive License Option..... 17
7.4.8.2 Nonexclusive Licenses..... 17
7.5 Non-Subject Inventions..... 18
7.5.1 Ownership of Non-Subject Inventions..... 18
7.5.2 Rights Under Other Agreements..... 18
7.5.3 No License to Non-Subject Inventions..... 18
7.6 Research License..... 18

Article 8. TANGIBLE PROPERTY..... 18

8.1 Title to Preexisting Tangible Property..... 18
8.2 Tangible Property Purchased by Collaborators to Perform Cooperative Work... 18
8.3 Title to Developed Tangible Property..... 19
8.4 Tangible Property Operational and Disposition Costs..... 19
8.5 Disposal of Tangible Property..... 19

Article 9. LIABILITY..... 19

9.1 Extent of Government Liability..... 19
9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Liability..... 19
9.3 *Force Majeure*..... 20

Article 10. GENERAL PROVISIONS..... 20

10.1 Characteristics of the Agreement..... 20
10.1.1 Entire Agreement..... 20
10.1.2 Severability..... 20
10.1.3 Interpretation of Headings..... 20
10.2 Agreement Between Collaborators..... 20
10.2.1 Governing Laws..... 20
10.2.2 Independent Parties/Entities..... 20
10.2.3 Assignment /Subcontracting..... 20

10.2.4 Disputes..... 21
10.2.4.1Settlement and Resolution..... 21
10.2.4.2Continuation of Cooperative Work..... 21
10.2.5 Waivers..... 21
10.2.6 Use of Name or Endorsements..... 22
10.3 Environment, Safety, and Health..... 22
10.4 U.S. Competitiveness..... 22
10.5 Public Release of This Agreement..... 22

Article 11. MODIFICATIONS AND NOTICES..... 22

11.1 Amendments..... 22
11.2 Termination..... 23
11.2.1 Termination by Mutual Consent..... 23
11.2.2 Unilateral Termination..... 23
11.3 Notices..... 23

Article 12. SURVIVING PROVISIONS..... 24

Article 13. DURATION..... 24

Article 14. SIGNATURES..... 24

Appendix A
Statement of Work

Appendix B
Confirmatory License Agreement

STANDARD
NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

PREAMBLE

Under authority of the U.S. Federal Technology Transfer Act of 1986 (Public Law 99-502, 20 October 1986, as amended), **NAVAIR ORLANDO TSD (Naval Air Warfare Center Training Systems Division)**, located at **12350 Research Parkway, Orlando, Florida 32826-3275**, and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, whose administrative offices are located at **12443 Research Parkway, Suite 207, Orlando, FL 32826-3252**, enter into this Cooperative Research and Development Agreement (CRADA), which shall be binding upon the Collaborators and their assignees according to the clauses and conditions hereof and for the term and duration set forth.

The U.S. Federal Technology Transfer Act of 1986, as amended, provides for making the expertise, capabilities, and technologies of U.S. Federal laboratories accessible to other Federal agencies; units of State or local government; industrial organizations (including corporations, partnerships and limited partnerships, and industrial development organizations); public and private foundations; nonprofit organizations (including universities); or other persons in order to improve the economic, environmental, and social well-being of the United States by stimulating utilization of U.S. Federally funded technology developments and/or capabilities.

NAVAIR ORLANDO TSD has extensive expertise, capabilities, and information in **Virtual Environments for Operator Training**, and, in accordance with the U.S. Federal Technology Transfer Act, desires to make this expertise and technology available for use in the public and private sectors.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory has the interest, resources, capabilities, and technical expertise to transition the results of Naval research and development for public use.

NOW THEREFORE, the Collaborators agree as follows.

Article 1. DEFINITIONS

As used in this Agreement, the following terms shall have the meanings defined below, which are equally applicable to both the singular and plural forms of nouns or any tense of verbs.

1.1 "Agreement" means this Cooperative Research and Development Agreement (CRADA) with its Appendices.

1.2 "Classified Information" means all Data classified in accordance with the national security laws of the United States.

1.3 "Collaborator" means the Navy participant or the Non-Navy participant represented and bound by the signatories of this Agreement.

1.4 "Controlled Unclassified Information (CUI)" means Government Data, Information, or materials provided to or resulting from this Agreement that may be export controlled, sensitive, for official use only, or otherwise protected by law, executive order, or regulation.

1.5 "Cooperative Work" means research, development, engineering, or other tasks performed under this Agreement by **NAVAIR ORLANDO TSD** or **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** working individually or together, pursuant to the Objectives (Article 2) and the Statement of Work (Appendix A).

1.6 "Data" means recorded information of any kind regardless of the form or method of the recording, including computer software.

1.7 "Effective Date" means the date of the last signature of the Collaborators executing this Agreement.

1.8 "Exclusive License" means the grant by the owner of Intellectual Property of the exclusive right to make, use, or sell a patented invention.

1.9 "Government" means the Government of the United States of America.

1.10 "Government Purpose Rights" means the right of the Government to use, duplicate, or disclose Data, in whole or in part, and in any manner, for Government purposes only, and to have or permit others to do so for Government purposes only. Government Purpose Rights includes competitive procurement, but does not include the right to have or permit others to use Data for commercial purposes.

1.11 "Information" means all data, trade secrets, and commercial and financial information. (Chapter 5 Subsection II of Title 5 USC)

1.12 "Intellectual Property" means the property of ideas, examples of which include, but are not limited to, patents, trademarks, copyrights, and trade secrets.

1.13 "Invention" means any invention or discovery that is or may be patentable or otherwise protected under Title 35, United States Code, or any novel variety of plant that is or may be patentable under the Plant Variety Protection Act. (15 USC 3703(9))

1.14 "Invention Disclosure" means the document identifying and describing to organizational management the Making of an Invention.

1.15 "Made" when used in conjunction with any Invention means the conception or first actual reduction to practice of such Invention. (15 USC 3703(10))

1.16 "Militarily Critical Technologies (MCT)" means those technologies identified in the Militarily Critical Technologies List and under the Export Administration Act of 1979, as amended.

1.17 "Non-Subject Data" means any Data that are not Subject Data.

1.18 "Non-Subject Invention" means any Invention that is not a Subject Invention.

1.19 "Patent Application" means an application for patent protection for an Invention with any domestic or foreign patent-issuing authority.

1.20 "Principal Investigator (PI)" means that person having the responsibility for the performance of the Cooperative Work on behalf of a Collaborator.

1.21 "Proprietary Information" means information that embodies trade secrets developed at private expense or business, commercial, or financial information that is privileged or confidential provided that such information:

is not known or available from other sources without obligations concerning its confidentiality;

has not been made available by the owners to others without obligation concerning its confidentiality;

is not already available to the Government without obligation concerning its confidentiality; and

has not been developed independently by persons who have had no access to the information. (FAR/DFARS Definition)

1.22 "Restricted Access Information" means Subject Data generated by **NAVAIR ORLANDO TSD** that would be Proprietary Information if the Information had been obtained from a non-Federal Collaborator participating in a CRADA (15 USC 3710a). Under 15 USC 3710a(c)(7)(B), the Collaborators mutually may agree to provide appropriate protection to Subject Data generated by **NAVAIR ORLANDO TSD** (Restricted Access Information) against public dissemination or release under the Freedom of Information Act (FOIA) for a period of up to five (5) years after development of the Information.

1.23 "Subject Data" means that Data first recorded in the performance of the Cooperative Work.

1.24 "Subject Invention" means any Invention Made in the performance of the Cooperative Work.

1.25 "Tangible Property" means personal or real property that can be physically touched or held.

1.26 "Unlimited Rights" means the right to use, modify, reproduce, release, disclose, perform, or display Data or Computer Programs in whole or in part, in any manner and for any purpose whatsoever, and to have or permit others to do so.

Article 2. OBJECTIVES

Background

NAVAIR ORLANDO TSD has been conducting Navy sponsored research in virtual environments for operator/small arms training for over 15 years. Results of this research have already been transitioned for many military training applications, and have also been the basis for technology transition in use by law enforcement. The Small Arms Simulation Training (SAST) laboratory at **NAVAIR ORLANDO TSD** now has in use an extensive range of software and data bases that have resulted from in-house **NAVAIR ORLANDO TSD** research work.

The University of Central Florida (UCF), with plans to establish a new laboratory for Human Factors research related to Homeland Defense, became aware of the work already done in **NAVAIR ORLANDO TSD's** SAST laboratory. Discussions were immediately begun to explore the mutual benefits of an agreement to work collaboratively.

The primary objective of this work will be to enhance mutual understand of operator performance under stress and develop possible applications of human factors research in this area for military training, and for homeland defense and related technologies

Benefit to Navy Partner: Research using **NAVAIR ORLANDO TSD** developed or enhanced source code may be expected to result in further enhancements to software and databases, at no expense to the government, and with direct use in military simulation-based training. Additionally, the anticipated UCF human factors research projects may be expected to have direct application and benefit for military training. Further, expertise will be provided to UCF on a cost basis, leveraging Navy research dollars.

Benefit to Non-Navy partner: **NAVAIR ORLANDO TSD** developed or enhanced source code for scenario generation, rural and urban databases, weapon tracking, image generation, networking/communication, etc., can provide the kind of virtual environment test bed required for planned human factors research, and can greatly accelerate work toward research objectives.

Article 3. RESPONSIBILITIES

The Collaborators shall provide personnel, facilities, and equipment necessary for, and shall perform, the Cooperative Work.

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities

The Cooperative Work done by **NAVAIR ORLANDO TSD** will be performed under the program guidance of **Ron Wolff, PI, NAVAIR ORLANDO TSD, AIR 4962**, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed

within the facilities of **NAVAIR ORLANDO TSD** or done on behalf of **NAVAIR ORLANDO TSD** by third parties in support of this Agreement.

UCF MURI / OPUS Laboratory personnel who perform Cooperative Work at **NAVAIR ORLANDO TSD** facilities will be supervised by the **UCF MURI / OPUS Laboratory** PI.

3.2 **UCF MURI / OPUS Laboratory** Personnel and Facilities

The Cooperative Work done by the **UCF MURI / OPUS Laboratory** will be performed under the program guidance of **Dr. Peter Hancock, UCF MURI / OPUS Laboratory** PI, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed within the facilities of the **UCF MURI / OPUS Laboratory** or done on behalf of the **UCF MURI / OPUS Laboratory** by third parties in support of this Agreement.

NAVAIR ORLANDO TSD personnel who perform Cooperative Work at the **UCF MURI / OPUS Laboratory** facilities will be supervised by **NAVAIR ORLANDO TSD** PI.

3.3 Security Regulations and Directives

Each Collaborator will abide by the safety and security regulations and directives of the host facility in which the Cooperative Work is being performed.

Article 4. REPRESENTATIONS AND WARRANTIES

4.1 **NAVAIR ORLANDO TSD**'s Representations and Warranties

NAVAIR ORLANDO TSD hereby warrants and represents to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** as follows:

4.1.1 **NAVAIR ORLANDO TSD** is a Federal laboratory of the U.S. Department of the Navy (Navy) as defined by 15 USC 3710a(d)(2)(A) and Department of Defense Instruction 5535.8, dated May 14, 1999.

4.1.2 The performance of the activities specified by this Agreement is consistent with the **Virtual Environments for Operator/Small Arms Training** and technology transfer missions of **NAVAIR ORLANDO TSD** (15 USC 3710a).

4.1.3 The Department of the Navy official executing this Agreement for **NAVAIR ORLANDO TSD** has the requisite power and authority to enter into this Agreement and to bind **NAVAIR ORLANDO TSD** to perform according to the terms of this Agreement.

4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Representations and Warranties

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory hereby warrants and represents to **NAVAIR ORLANDO TSD** as follows:

4.2.1 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not directly or indirectly controlled by a foreign company or government (Executive Order 12591, Section 4 (a)). **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, as of the Effective Date of this Agreement, is a part of a **university** duly organized, validly existing, and in good standing under the laws of **the State of Florida.**

4.2.2 The official executing this Agreement for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** has the requisite power and authority to enter into this Agreement and to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to perform according to the terms of this Agreement.

4.2.3 The University of Central Florida and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** have taken all actions required by law, or otherwise, to authorize the execution and delivery of agreements, such as this Agreement.

4.2.4 The execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any agreement binding on **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. Furthermore, the execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any valid order of any court, or any regulatory agency or other body having authority to which **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is subject.

4.2.5 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not presently subject to debarment or suspension by any agency of the Government. Should **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** be debarred or suspended during the term of this Agreement or thereafter, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** will notify **NAVAIR ORLANDO TSD** within thirty (30) days of receipt of a final notice. **NAVAIR ORLANDO TSD** may then elect to terminate this Agreement and any licenses and options granted under this Agreement.

4.2.6 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is not** a small business as defined in 15 USC 632 and implementing regulations (13 CFR 121.101 et seq.) of the Administrator of the Small Business Administration.

4.3 Joint Representations

The Collaborators make the following representations.

4.3.1 There is no express or implied warranty as to any research, Invention, or product, whether tangible or intangible. In particular, the Collaborators make no express or implied warranty as to the merchantability or fitness for a particular purpose of any research, Invention, or product, whether tangible or intangible. Likewise, the Collaborators make no express or implied warranty as to any Cooperative Work, Subject Invention, Subject Data, or other product resulting from the Cooperative Work.

4.3.2 The use and dissemination of Information and materials exchanged under this Agreement will be in accordance with all U.S. laws and regulations, including those pertaining to national security and export control. Nothing in this Agreement shall be construed as a license to export Information or to permit any disclosure in violation of law, regulation, or Department of Defense policy. The exporting Collaborator is responsible for obtaining any export licenses that may be required by U.S. Federal law.

Article 5. FUNDING

5.1 Payment Schedule

Consistent with the information in Appendix A, **NAVAIR ORLANDO TSD** will provide source code/databases developed or enhanced by **NAVAIR ORLANDO TSD** to the **MURI / OPUS** lab at no cost to **UCF**. If, at a later date during the period of this CRADA, there is a mutually determined need for additional expertise from **NAVAIR ORLANDO TSD**, then the following will apply. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees to pay **NAVAIR ORLANDO TSD** on a cost basis, as mutually determined to be required, during the period of this CRADA, as specified in Appendix A.

Checks will be payable to:

NAVAIR ORLANDO TSD, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments

Each check and its cover correspondence shall refer to Navy CRADA number "NCRADA-**NAVAIR ORLANDO TSD-02-037**."

Checks will be mailed to:

NAVAIR ORLANDO TSD, 12350 Research Parkway (AIR 4.9T Nimmo), Orlando, Florida 32826-3275, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments.

5.2 Insufficient and Excess Funds

NAVAIR ORLANDO TSD may discontinue performance under this Agreement if the funds provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** for performance by **NAVAIR ORLANDO TSD** are insufficient or are not provided as specified in Article 5.1. In the event **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to tender the Government the required payment within fifteen (15) days after its respective due date, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall be in default under this Agreement for failure to make payments. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on**

behalf of the MURI / OPUS Laboratory is in default for this reason, NAVAIR ORLANDO TSD shall notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** does not cure the default within fifteen (15) days of mailing date of notice, **NAVAIR ORLANDO TSD** may proceed to terminate the Agreement in accordance with Article 11.2.2, may cancel any option for an Exclusive License to a Subject Invention, and may terminate any Exclusive License granted pursuant to this Agreement.

Funds that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** paid under Article 5.1 and that **NAVAIR ORLANDO TSD** has not obligated or expended at the time of completion, expiration, or termination of this Agreement shall be returned to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** after **NAVAIR ORLANDO TSD's** submission of a final fiscal report to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**.

5.3 No New Commitments

NAVAIR ORLANDO TSD shall make no new commitments concerning this Agreement after receipt of a written termination notice from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in accordance with Article 11.2 and shall, to the extent practicable, cancel all outstanding commitments by the termination date. Should such cancellation result in any costs incurred by **NAVAIR ORLANDO TSD, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that such costs shall be chargeable against any funding that it provided to **NAVAIR ORLANDO TSD.**

5.4 Accounting Records

NAVAIR ORLANDO TSD shall maintain current accounts, records, and other evidence supporting all its expenditures against funding provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement and shall retain such records for at least twelve (12) months after the completion, expiration, or termination of this Agreement. **NAVAIR ORLANDO TSD** shall provide **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** a financial report within four (4) months after completion, expiration, or termination of this Agreement.

Article 6. REPORTS AND PUBLICATIONS

6.1 Final Report

The Collaborators shall submit to each other a final report within four (4) months of the completion, termination, or expiration of this Agreement that includes the results obtained and a list of all Subject Inventions Made.

6.2 Agreement to Confer Prior to Publication or Public Disclosure

The Collaborators agree to confer and consult prior to any publication or public disclosure of Subject Data to ensure that no Proprietary Information, Restricted Access Information, Government Classified Information, CUI, or MCT Information is released and that patent rights are not compromised. Prior to any such publication or public disclosure of Subject Data, each Collaborator shall be offered a period not less than fifteen (15) days and not to exceed thirty (30) days, unless otherwise mutually agreed in writing by the Collaborators, to review any proposed abstract, publication, presentation, or other document for public disclosure that contains Subject Data. For the purposes of this Article, the term "disclosure" shall include, but not be limited to, submission of any manuscript for peer review prior to publication. It is the responsibility of the Collaborator intending to make public disclosure of Subject Data to notify the other Collaborator of such intent.

If a Collaborator objects to a proposed public disclosure, that Collaborator must so notify the other Collaborator within thirty (30) days of the date of notice of intent to disclose publicly. If no objection is received by the Collaborator intending to make public disclosure, concurrence is assumed. If a Collaborator objects on the grounds that patent rights may be compromised, a Patent Application must be filed by the responsible Collaborator within ninety (90) days of the date of notification of intent to make public disclosure, or by another date mutually agreed to by the Collaborators. If a Collaborator objects to the release of Information on the grounds that the Information is Proprietary Information, Restricted Access Information, or Information whose dissemination is restricted by U.S. security laws or regulations, the disclosure shall be postponed until the Information no longer meets the definitions of Proprietary Information, Restricted Access Information, or is no longer covered by U.S. security laws or regulations.

6.3 Classified Information

Any presentation that includes Subject Data that are Classified Information or otherwise restricted Data must have prior review and approval by **NAVAIR ORLANDO TSD** pursuant to the pertinent security laws, regulations, and directives.

Article 7. INTELLECTUAL PROPERTY

7.1 Data

7.1.1 General Provisions Applying to All Data

7.1.1.1 Ownership

Each Collaborator shall have title to all Data generated by that Collaborator.

7.1.1.2 No Implied License

Unless otherwise specifically provided, the Collaborators agree that the exchange of Data of any kind does not confer a license to any Invention claimed in any

patent or Patent Application or to the subject matter of any copyright, trademark/service mark, or other form of Intellectual Property protection.

7.1.1.3 Marking of Data

7.1.1.3.1 Data Provided With Less Than Unlimited Rights

Each Collaborator shall mark all Data that it provides with less than Unlimited Rights with a marking that clearly identifies the limited rights.

7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall place a proper proprietary marking on each medium used for recording Data that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** delivers to **NAVAIR ORLANDO TSD** under this Agreement that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** asserts is Proprietary Information. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall request in writing if it wishes Subject Data generated by **NAVAIR ORLANDO TSD** to be marked as Restricted Access Information. The Collaborators together shall confer to determine if such marking is appropriate, with reference to the Definitions of Article 1. If the Collaborators mutually agree to the marking then:

(a) For Non-Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory – NAVAIR ORLANDO TSD** MAY USE ONLY FOR PURPOSE OF CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(b) For Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** – GOVERNMENT HAS GOVERNMENT PURPOSE RIGHTS UNDER CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(c) For Data that are Restricted Access Information, the marking shall read:

"RESTRICTED ACCESS INFORMATION – PROTECT IN ACCORDANCE WITH CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**" UNTIL **SEPTEMBER 2005**".

7.1.1.3.3 Data That are Subject to 35 USC 205

NAVAIR ORLANDO TSD shall mark Data it provides under this Agreement that disclose one or more Inventions in which the Government owns or may own a right, title or interest, and that are subject to confidentiality under 35 USC 205. Such Data shall be marked:

"**NAVAIR ORLANDO TSD** DATA PROTECTED FROM RELEASE OR DISCLOSURE UNDER 35 USC 205."

7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted

Each Collaborator shall mark all Data that are Classified Information, CUI, MCT, or otherwise restricted by U.S. security or export control laws or regulations that it provides under this Agreement.

7.1.1.4 Protection of Data

Except for the rights granted in Article 7.1.2.2, Data shall be protected in accordance with the proper markings of its owner and as provided by, at a minimum, the requirements of 15 USC 3710a. Proprietary Information will be protected only if it is properly marked as such. Information provided in intangible form that is Proprietary Information must be designated Proprietary Information at the time it is delivered, followed within fifteen (15) days by a writing summarizing the exact Information to be protected. The Collaborator receiving Information in an intangible form that is designated as Proprietary Information shall be responsible for protecting the Information as Proprietary Information during the fifteen (15) day notification period. After the fifteen (15) day period, if no written summary has been received, the receiving Collaborator need not continue to protect the Information received in intangible form.

Restricted Access Information shall be protected from public dissemination for up to five (5) years, as mutually agreed.

Classified Information, CUI, MCT, or otherwise restricted Information shall be protected in accordance with the security laws of the United States.

7.1.1.5 Release of Data Under the Freedom of Information Act

Data in the possession of **NAVAIR ORLANDO TSD** that are not marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or Restricted Access Information must be released by **NAVAIR ORLANDO TSD** where such release is required pursuant to a request under the Freedom of Information Act (FOIA) (5 USC 552). **NAVAIR ORLANDO TSD** shall protect Data that are properly marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI**

/ **OPUS Laboratory** or Restricted Access Information from release under the FOIA for as long as the marked Data meet the definition of CUI, Proprietary Information or Restricted Access Information. Prior to release of any such Data, **NAVAIR ORLANDO TSD** shall promptly notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** of any request for Data of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** regardless of whether the requested Data are marked Proprietary Information or as required under the public records law.

7.1.2 Subject Data

7.1.2.1 Delivery of Requested Subject Data

Each Collaborator shall have the right to review and receive delivery of all Subject Data generated by the other Collaborator. Requested Subject Data shall be delivered to the requesting Collaborator within fifteen (15) days of the request.

7.1.2.2 Rights in Subject Data

Except as represented in Article 4.3.2, the Collaborators shall have Unlimited Rights in all Subject Data that are not Proprietary Information or Restricted Access Information. Notwithstanding 15 USC 3710a, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** grants Government Purpose Rights in any Subject Data furnished by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to **NAVAIR ORLANDO TSD** under this Agreement that are properly marked as Proprietary Information. The Government has Government Purpose Rights in Subject Data that are Restricted Access Information.

7.1.3 Rights in Non-Subject Data

The Collaborators shall have Unlimited Rights in any Non-Subject Data delivered under this Agreement that are not Proprietary Information.

NAVAIR ORLANDO TSD has a limited right to use, reproduce, and disclose only to Government employees for use in support of the Cooperative Work any Non-Subject Data that are properly marked as Proprietary Information and are provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement. Such Proprietary Information can be used only for the purpose of performing the Cooperative Work unless consent to other use or disclosure is obtained from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in writing.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall have a limited right to use, reproduce, or disclose Non-Subject Data that may describe one or more Inventions in which the Government owns or may own a right, title or interest, if such Non-Subject Data are provided by **NAVAIR ORLANDO TSD** under this Agreement. In accordance with 35 USC 205, such Non-Subject Data are to be held in confidence. Such Non-Subject Data shall be properly marked by **NAVAIR ORLANDO TSD** and the limited rights of **The UCF Office of Research, UCF on behalf of its**

Board of Trustees on behalf of the MURI / OPUS Laboratory shall be defined by a separate non-disclosure agreement.

7.2 Copyrights

7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory may copyright works of authorship prepared pursuant to this Agreement if eligible for copyright protection under Title 17 USC.

7.2.2 Copyright License to the Government

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory grants to the Government a nonexclusive, irrevocable, paid-up license in copyrighted works of authorship, including software (17 USC 106) prepared pursuant to this Agreement for any purpose, consistent with the rights in Data described in Article 7.1.

7.2.3 Copyright Statement

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall include the following statement on any text, drawing, mask work or other work of authorship, that may be copyrighted under 17 USC, that is created in the performance of this Agreement:

"The U.S. Government has a copyright license in this work pursuant to a Cooperative Research and Development Agreement with **NAVAIR ORLANDO TSD.**"

7.3 Trademarks and Service Marks

7.3.1 Ownership of Trademarks and Service Marks

The Collaborator first establishing a trademark or service mark for goods or services with which the mark is used shall be considered the owner of the mark.

7.3.2 Obligation of Employees to Report Trademarks and Service Marks

Employees of both Collaborators shall report the adoption of a trademark or service mark associated with the Cooperative Work to their employer within thirty (30) days of the first use of the mark. Use includes internal use of any product or service of the Cooperative Work.

7.3.3 Obligation of Collaborators to Notify Each Other

Each Collaborator shall notify the other Collaborator within thirty (30) days of their employee's report of the first use of a trademark or service mark.

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

Each Collaborator shall be entitled to own the Subject Inventions of its employees. Each Collaborator shall cooperate with the other Collaborator to obtain inventor signatures on Patent Applications, assignments or other documents required to secure Intellectual Property protection. For any Invention Made jointly by employees of the Collaborators, each Collaborator shall have ownership of the Subject Invention in the form of an undivided interest.

7.4.4 Filing of Patent Applications

7.4.4.1 Filing of Patent Applications on Solely Made Inventions

Each Collaborator has primary responsibility for filing Patent Applications on the Subject Inventions of its employee(s).

Notwithstanding such primary responsibility, by mutual agreement, the Collaborators may identify which Collaborator shall file a Patent Application on any Subject Invention.

7.4.4.2 Filing of Patent Applications on Jointly Made Inventions

In the case of an Invention jointly Made by employees of both Collaborators, the Collaborators shall confer and agree as to which Collaborator will file any Patent Application. Officers of the non-filing Collaborator shall cooperate with the filing Collaborator to obtain signatures on documents that are needed to file a Patent Application.

7.4.4.3 Preserving Intellectual Property Rights

The Collaborator responsible for filing of a Patent Application on any Subject Invention shall file such Patent Application at least sixty (60) days prior to any bar date or one year from the date the Invention Disclosure was received, whichever comes first. If no Patent Application is filed within the specified time period, the other Collaborator may assume control of filing the Patent Application and take title to the Subject Invention on ten (10) days written notification. The Collaborator that relinquished the responsibility to file shall retain a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Subject Invention practiced throughout the world by or on its behalf.

7.4.4.4. Filing Deadlines

The Collaborator responsible for filing any Patent Application for a Subject Invention shall notify the other Collaborator of all filing deadlines for prosecution of any Patent Application and maintenance of any patents on the Subject Invention. Notwithstanding the primary responsibility defined in Article 7.4.4.1, sixty (60) days prior to any filing deadline, the Collaborators shall confer to determine if the filing Collaborator intends to respond to the filing deadline. The non-filing Collaborator will be permitted to take action if the filing Collaborator declines.

7.4.4.5 Copies and Inspection

7.4.4.5.1 Copies of Prosecution Papers

Each Collaborator filing a Patent Application on a Subject Invention shall provide the other Collaborator with a copy of any communication relating

to prosecution of said Patent Application within thirty (30) days of receipt of such communication.

7.4.4.5.2 Access to Patent Application File and Right to Make Copies

Upon written request, the filing Collaborator shall give the other Collaborator an Associate Power of Attorney, with authorization to access the Patent Application, make copies, and, in the event the filing Collaborator fails or declines to take action, do all that is necessary to secure Intellectual Property protection for the Subject Invention.

7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application

In the event both Collaborators decline to file a Patent Application on a Subject Invention, the Government will renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

In the event both Collaborators decline to file a Patent Application on a Subject Invention, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** may, at its sole discretion, renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

7.4.5 Nonexclusive License to Subject Inventions

7.4.5.1 Nonexclusive License Grant

Each Collaborator grants to the other Collaborator a nonexclusive, irrevocable, paid-up license to practice a Subject Invention Made by employees of the granting Collaborator or have the Subject Invention practiced throughout the world by or on behalf of the other Collaborator. No nonexclusive license granted under this Agreement shall permit licensee to grant sublicenses.

7.4.5.2 Confirmatory Nonexclusive License Agreement

Each Collaborator has the obligation to provide a Confirmatory License Agreement (Appendix B) to the other Collaborator for each nonexclusive license within ninety (90) days of the date of filing.

7.4.6 Option for Exclusive License to Subject Inventions

NAVAIR ORLANDO TSD gives **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** the option of acquiring an Exclusive License for the field of use **(virtual environments for operator/small arms training)** in the Government's rights in any Subject Invention Made in whole or in part by a **NAVAIR ORLANDO TSD** employee. The license shall be for reasonable consideration. In

order to exercise this option, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must notify **NAVAIR ORLANDO TSD** in writing within one hundred and eighty (180) days of the filing of a Patent Application. Unless another time period is mutually agreed upon between the Collaborators, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must execute an Exclusive License to the Subject Invention within one hundred and eighty (180) days of election to exercise the option, or the Invention shall be made available for licensing by the public in accordance with 37 CFR Part 404.

Any Exclusive License granted by the Government in a Subject Invention is subject to the statutorily required reservation by the Government of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the Government (15 USC 3710a).

7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement

No license granted under this Agreement shall be assigned, licensed or otherwise disposed of except to the successor in interest of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which such license pertains.

7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions

7.4.8.1 Exclusive Licenses and Exclusive License Option

NAVAIR ORLANDO TSD may terminate any Exclusive License or cancel any option for an Exclusive License to a Subject Invention granted under this Agreement in the event that:

(a) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is in default for failure to make payment as agreed in Article 5; or

(b) The Agreement is terminated unilaterally by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**; or

(c) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to perform according to the Statement of Work (Appendix A); or

(d) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** becomes a foreign owned, controlled, or influenced (FOCI) organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.4.8.2 Nonexclusive Licenses

NAVAIR ORLANDO TSD shall terminate any nonexclusive license to a Subject Invention granted under this Agreement if **The UCF Office of Research,**

UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory becomes a FOCI organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.5 Non-Subject Inventions

7.5.1 Ownership of Non-Subject Inventions

Each Collaborator owns its Non-Subject Inventions.

7.5.2 Rights Under Other Agreements

Nothing in this Agreement is intended to change the rights in Intellectual Property acquired by the Collaborators in any other contract or agreement between the **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** and the Government.

7.5.3 No License to Non-Subject Inventions

This Agreement does not grant any Collaborator a license, express or implied, to any Non-Subject Invention.

7.6 Research License

Each Collaborator shall allow the other Collaborator to practice any of its Non-Subject Inventions for the purpose of performing the Cooperative Work.

No license, express or implied, for commercial application(s) is granted to either Collaborator in Non-Subject Inventions by performing the Cooperative Work.

For commercial application(s) of Non-Subject Inventions, a license must be obtained from the owner.

Article 8. TANGIBLE PROPERTY

8.1 Title to Preexisting Tangible Property

Each Collaborator shall retain title to all Tangible Property to which it had title prior to the Effective Date of this Agreement.

8.2 Tangible Property Purchased by Collaborators to Perform the Cooperative Work

Each Collaborator shall retain title to all Tangible Property that it purchases during the period of this Agreement. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** cannot take title to any Government Tangible Property under this Agreement. Collaborator consumables to be used in the Cooperative Work of this Agreement are the property of the purchasing Collaborator until consumed.

8.3 Title to Developed Tangible Property

All Tangible Property developed under this Agreement with all components purchased by one Collaborator shall be the property of that Collaborator. Tangible Property having any component purchased by **NAVAIR ORLANDO TSD** shall be the property of the Government, unless such Tangible Property can reasonably be separated without damage to the other individual components. After this Agreement is completed, expired, or terminated, if separation of components can be made without damage, the Collaborators may, by mutual agreement, separate the Tangible Property into its components and the separated components shall remain the property of the Collaborator that purchased them.

8.4 Tangible Property Operational and Disposition Costs

During the period of and upon completion, expiration, or termination of this Agreement, each Collaborator shall be responsible for all costs of maintenance, removal, storage, repair, disposal, and shipping of all Tangible Property to which it has title.

8.5 Disposal of Tangible Property

Disposal of Tangible Property shall be in accordance with applicable U.S. Federal, State, and local property disposal laws, environmental laws, and regulations.

Article 9. LIABILITY

9.1 Extent of Government Liability

The Government shall be solely liable for the negligent or wrongful acts of its officers and employees to the extent provided for in the Federal Tort Claims Act (28 USC 2671 et. seq.) and in other applicable laws and regulations of the United States that specifically waive sovereign immunity. Nothing in this Agreement shall be construed as a waiver of the sovereign immunity of the United States.

9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Liability

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is solely responsible for its actions and the actions of those acting for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in the performance of this Agreement and for any damages that may arise from any suit, action, or claim, and for any costs from or incidental to any suit, action, or claim, including but not limited to settlement and defense costs. Further, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that in any suit, action or claim brought by anyone not a party to this Agreement based on actions of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not pursue any actions to enter the Government as a party in such suit, action or claim unless the Government has some liability under the Federal Tort Claims Act.

9.3 *Force Majeure*

No Collaborator shall be liable for the consequences of any *force majeure* that (1) is beyond its reasonable control; (2) is not caused by the fault or negligence of such Collaborator; (3) causes such Collaborator to be unable to perform its obligations under this Agreement; and (4) cannot be overcome by the exercise of due diligence. In the event of the occurrence of a *force majeure*, the Collaborator unable to perform shall promptly notify the other Collaborator. The Collaborators shall suspend performance only for such period of time as is necessary to overcome the result(s) of the *force majeure* and shall use their best efforts to resume performance as quickly as possible.

Article 10. GENERAL PROVISIONS

10.1 Characteristics of the Agreement

10.1.1 Entire Agreement

This Agreement constitutes the entire agreement between the Collaborators concerning the Cooperative Work and supersedes any prior understanding or written or oral agreement relative to the Cooperative Work.

10.1.2 Severability

The illegality or invalidity of any Article of this Agreement shall not impair, affect, or invalidate any other Article of this Agreement.

10.1.3 Interpretation of Headings

Headings of the Articles of this Agreement are for convenience of reference only and do not form a part of this Agreement and shall in no way affect the interpretation thereof.

10.2 Agreements Between Collaborators

10.2.1 Governing Laws

United States Federal Laws shall govern this Agreement for all purposes.

10.2.2 Independent Parties/Entities

The relationship of the Collaborators to this Agreement is that of independent parties and not as agents of each other, partners, or participants in a joint venture. Each Collaborator shall maintain sole and exclusive control over its personnel and operations.

10.2.3 Assignment/Subcontracting

10.2.3.1 Neither Collaborator may allow third parties to perform any part of the Cooperative Work under this Agreement without express written consent of the other

Collaborator. If consent is obtained, the Collaborator requesting such consent shall remain fully responsible for the portion of the Cooperative Work to be accomplished under a third-party agreement, and the third party is not a Collaborator of this Agreement. Any third-party agreement to perform a portion of the Cooperative Work shall contain terms consistent with this Agreement.

10.2.3.2 This Agreement shall not be assigned or otherwise transferred by either Collaborator without the prior written consent of the other Collaborator, except to the successor of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which this Agreement pertains.

10.2.3.3 If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee is a U.S. company, and becomes, during the term of this Agreement or thereafter, directly or indirectly owned, controlled, or influenced by a foreign company or government (FOCI), then **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee shall promptly notify **NAVAIR ORLANDO TSD** to that effect.

10.2.4 Disputes

10.2.4.1 Settlement and Resolution

NAVAIR ORLANDO TSD and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agree to use reasonable efforts to reach a fair settlement of any dispute. If such efforts are unsuccessful, remaining issues in dispute will be referred to the signatories or their successors for resolution. If a dispute continues, the remaining issues may be submitted to the Chief of Naval Research (CNR), or the CNR designee, for resolution. This Agreement does not prevent any Collaborator from pursuing disputes in a U.S. Federal court of competent jurisdiction. No Collaborator will pursue litigation in a U.S. Federal court until after the CNR, or the CNR designee, decides the dispute, or until sixty (60) days after the dispute was first submitted to the CNR, or the CNR designee, whichever comes first.

10.2.4.2 Continuation of Cooperative Work

If payments or installment payments are to be made as stated under Article 5, **NAVAIR ORLANDO TSD** will not start or continue cooperative work until payments or installment payments are received.

10.2.5 Waivers

None of the provisions of this Agreement shall be considered waived by either Collaborator unless such waiver is given in writing to the other Collaborator, signed by the executing official of this Agreement or the official's successor having the authority to bind the Collaborator making the waiver. The failure of either Collaborator to insist upon strict performance of any of the terms and conditions herein, or failure or delay to exercise any rights provided herein or by law shall not be deemed a waiver of any right of either Collaborator under this Agreement.

10.2.6 Use of Name or Endorsements

Except as provided for in Article 7.2.3, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not use the name of **NAVAIR ORLANDO TSD** or any other Government entity on any product or service that is directly or indirectly related to either this Agreement or any patent license or assignment associated with this Agreement without the prior approval of **NAVAIR ORLANDO TSD**. By entering into this Agreement, **NAVAIR ORLANDO TSD** does not directly or indirectly endorse any product or service provided, or to be provided, by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, its successors, assignees, or licensees. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not in any way imply that the Department of the Navy endorses any such product or service.

10.3 Environment, Safety, and Health

Each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste in its custody during the course of this Agreement. At the conclusion of this Agreement, each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste still in its possession. Each Collaborator shall obtain at its own expense all necessary permits and licenses as required by U.S. Federal, State, and local law and shall conduct such handling, control, and disposition in a lawful and environmentally responsible manner. Each Collaborator is responsible for all required environmental, safety, and health compliance, notice, and monitoring related to its facility in accordance with U.S. Federal, State, and local law and regulations. Collaborators shall abide by the environmental, safety, and health directives of the host facility in which the Cooperative Work is being performed, and any U.S. Federal, State, or local laws and regulations pertaining to environment, safety, and health that are applicable to the host facility.

10.4 U.S. Competitiveness

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory agrees that any product, process, or service using Intellectual Property arising from the performance of this Agreement shall be manufactured substantially in the United States.

10.5 Public Release of This Agreement

This Agreement, without funding information (Article 5) and Appendices, may be released to the public.

Article 11. MODIFICATIONS AND NOTICES

11.1 Amendments

If a Collaborator wishes to modify this Agreement, the Collaborators shall confer in good faith to determine the desirability of such modification. Such modification shall not be effective until a written amendment is signed by both executing officials of this Agreement or their successors.

11.2 Termination

11.2.1 Termination by Mutual Consent

The Collaborators may elect to terminate this Agreement at any time by mutual consent. Such termination shall not be effective until a written termination agreement is signed by both executing officials of this Agreement or their successors.

11.2.2 Unilateral Termination

A Collaborator may unilaterally terminate this entire Agreement at any time by giving the other Collaborator written notice signed by the executing official of this Agreement or his/her successor, not less than thirty (30) days prior to the desired termination date. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** unilaterally terminates this Agreement, any option for an Exclusive License to a Subject Invention and any Exclusive License to a Subject Invention granted by or pursuant to this Agreement shall simultaneously be terminated.

11.3 Notices

All notices pertaining to or required by Articles of this Agreement, except those pertaining solely to the prosecution of any patent, trademark, or service mark, shall be in writing and shall be signed by an authorized representative of the Technology Transfer Office for **NAVAIR ORLANDO TSD** or the preferred contact for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, and all such notices shall be delivered by hand, sent by courier with proper registration, or sent by certified mail, return receipt requested, with postage prepaid, addressed as follows:

If to **NAVAIR ORLANDO TSD**:

Ms. Patricia R. Nimmo
Technology Transfer Project Manager
NAVAIR ORLANDO TSD (AIR 49T)
12350 Research Parkway
Orlando, FL 32826-3275

If to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

Ms. Kim Smith
Senior Contract Specialist
Office of Research
University of Central Florida
Orlando Tech Center
12443 Research Parkway, Suite 207
Orlando, Fl 32826-3252

A Collaborator shall notify the other Collaborator of a change of address in the manner set forth above.

Notices pertaining solely to the prosecution of any patent, trademark, or service mark related to this Agreement shall be in writing and shall be signed by and sent to the Collaborator's legal counsel for Intellectual Property. Legal counsel for Intellectual Property for each Collaborator shall send a copy of any such notice to the Technology Transfer Office for he **NAVAIR ORLANDO TSD**. If either Collaborator fails to identify such counsel upon request, then such notices shall be sent to the points of contact specified above.

Article 12. SURVIVING PROVISIONS

The Articles covering Definitions, Representations and Warranties, Funding, Reports and Publications, Intellectual Property, Tangible Property, Liability, General Provisions, Modifications and Notices, and Surviving Provisions shall survive the completion, termination, or expiration of this Agreement.

Article 13. DURATION

This Agreement expires three years after its Effective Date, unless otherwise extended in writing according to the provisions of Article 11.

Article 14. SIGNATURES

For **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

I, the undersigned, am duly authorized to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002

By: M.J. Soileau

Title: Vice President, Office of Research

For the Department of the Navy:

I, the undersigned, by 15 USC 3710a and Navy regulations, am duly authorized to bind the U.S. Navy to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002.

By:

R. A. MOHLER

Title: Captain, U.S. Navy
Commanding Officer

Navy Organization: NAVAIR ORLANDO, TRAINING SYSTEMS DIVISION

This Page Intentionally Left Blank

APPENDIX A

STATEMENT OF WORK
BETWEEN
NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND
THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

NAVAIR ORLANDO TSD will:

1. Provide, for use in the UCF MURI/OPUS laboratory, source code/databases developed or enhanced by NAVAIR ORLANDO TSD as part of Navy sponsored research and development, e.g., Host Software, Image Generator Software, Weapon Tracking Software, Networking/Communications Software, Data Collection/Analysis Software, Scenario Builder Software, Rural and Urban Databases

2. Provide expertise, on a cost basis and as mutually determined to be required, on utilization of government provided software and databases for UCF MURI/OPUS Laboratory research

3. Provide, on a cost basis and as mutually determined to be required, expertise for further hardware/software/database development, testing, and training applications

4. Participate, on a cost basis, in mutually selected specific research projects conducted under the auspices of the UCF MURI/OPUS Laboratory

5. Provide access to NAVAIR ORLANDO TSD customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

3. Mutually develop processes separate from this agreement to allow exchange of funds, should that be mutually determined to be required during the period of this CRADA

NOTE 1: The parties agree that Article 9.2 Extent of Liability, will be subject to and interpreted in accordance with Florida Statutes, Section 768.28.

NOTE 2: FOREIGN ACCESS TO TECHNOLOGY

All work described in this SOW shall be subject to the following with regard to foreign access to technology.

The Collaborators agree that research findings and technology developments under this CRADA may constitute a significant enhancement to the national defense. Accordingly, access to important technology developments under this CRADA by foreign firms or institutions must be carefully controlled. These controls are in addition to, and are not intended to change or supercede, the provisions of the International Traffic in Arms Regulation (22 CFR pt. 121 et seq.) the DoD Industrial Security Regulation (DoD 5220.22-R) and the Department of Commerce Export Regulation (15 CFR pt.770 et seq.)

The non-Navy Collaborator shall provide timely notice to the government's designated Principal Investigator of any proposed transfer of research findings and technology developments under this CRADA to foreign firms or institutions. If the government determines that the transfer may have adverse consequences to the national security interests of the United States, the non-Navy Collaborator and the government shall jointly endeavor to find alternatives to the proposed transfer which obviate or mitigate potential adverse consequences of the transfer but which provide substantially equivalent benefits to the non-Navy Collaborator.

APPENDIX B

CONFIRMATORY LICENSE AGREEMENT	*1. APPLICATION FOR (Title of Invention)*
2. INVENTOR(S) AND AFFILIATION	
3. PATENT APPLICATION SERIAL NO.	*4. PATENT APPLICATION FILING DATE*
5. NAVY ACTIVITY (Name, address, point of contact)	*6. NON-NAVY ACTIVITY (Name, address, point of contact)*
7. CRADA AGREEMENT NO.	*8. DATE OF THIS AGREEMENT*

9. The Invention identified above is a "Subject Invention" under Article 7 Intellectual Property included with the CRADA identified in Box 7 between the Department of the Navy and Non-Navy Activity identified in Box 6.

This document is confirmatory of the nonexclusive, irrevocable, paid-up license to practice the identified Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the receiving party, and of all other rights acquired by the receiving party by the referenced clause.

This license is granted to

_____ *the Government*

(Select one)

_____ *Non-Navy Activity identified in Box 6*

under this CRADA in the identified Invention, Patent Application and any resulting patent.

The licensee is hereby granted an irrevocable power to inspect and make copies of the above-identified Patent Application.

_____ *ACTIVITY NAME OF LICENSOR*

_____ *SIGNATURE*

_____ *NAME (Typed or Printed)*

_____ *TITLE*

_____ *BUSINESS TELEPHONE*

University of Central Florida

Multi-Disciplinary University Research Initiative

MURI-OPUS

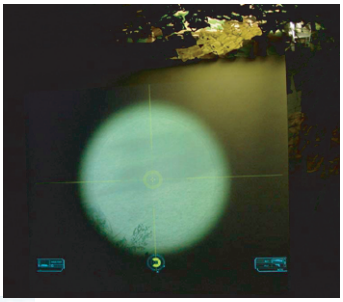
Operator Performance Under Stress

www.mit.ucf.edu

VOLUME ONE







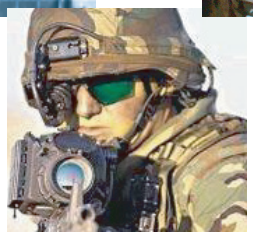




Table of Contents

NCRADA

A Collaborative Partnership between:

University of Central Florida

Contact Information:
University of Central Florida • MIT² Laboratory
3100 Technology Pkwy, Suite 337 • Partnership II • Orlando, Florida 32826-0544
http://www.mit.ucf.edu

Table of Contents

MURI Divisions

University of Central Florida (Peter Hancock)

Supported Work

A Descriptive Framework for the Evaluation of Stress Effects on Operator Performance *Poster Presented* 45th Annual Human Factors and Ergonomics Society *Hancock, P.A., Ganey, H.C.N., Mouloua, M., Salas, E., Gilson, R., Greenwood-Erickson, A., Parasuraman, R., Harris, W., Leon, A. & Smith, K. (2001)*

A Descriptive Framework for the Evaluation of Stress Effects on Operator Performance *Proceedings* 45th Annual Human Factors and Ergonomics Society *Hancock, P.A., Ganey, H.C.N., Mouloua, M., Salas, E., Gilson, R. & Greenwood-Erickson, A. (2001)*

Facilitated Work

Ergonomics as a Foundation for a Science of Purpose *Refereed Publication* Hancock, P.A. & Diaz, D.D. (2001) Ergonomics as a foundation for a science of purpose. *Theoretical Issues in Ergonomic Science,* **3** (2), 115-123.

Catholic University (Raja Parasuraman)

Supported Work

Quarterly Update 12/2001

University of Minnesota (Wayne Harris)

Supported Work

Phase I – Progress report 08/2001 *Leon, A.J.*

Phase I – Progress report 09/2001 *Leon, A.J.*

Phase I – Progress report 12/2001 *Leon, A.J.*

Kansas State (Kip Smith)

Supported Work

Progress report - Smith, K. 08/2001

Progress report - Smith, K. 09/2001

A DESCRIPTIVE FRAMEWORK FOR THE EVALUATION OF STRESS EFFECTS ON OPERATOR PERFORMANCE

P.A. Hancock[1], H.C.N. Ganey[1], M. Mouloua[1], E. Salas[1], R. Gilson[1], A. Greenwood-Ericksen[1],
R. Parasuraman[2], W. Harris[3], A. Leon[3], *and* K. Smith[4]

[1]*University of Central Florida*
[2]*Catholic University of America*
[3]*University of Minnesota*
[4]*Kansas State University*
URL: http://mit.ucf.edu/muri EMAIL: hganey@mail.ucf.edu

ABSTRACT

This work provides a general descriptive framework that relates the action of stress to change in operator performance capacity. The key advantage of our approach is that it allows us to capture simultaneously the influences of both physical and cognitive forms of stress and their singular and interactive effects on response efficiency. This is accomplished within the model since we propose that response processes to physiological challenge and psychological challenge are fundamentally identical in their mode of operation. By considering the commonalities of brain function with the response processes of other organs of the body, we can now use the extensive existing body of physiological insight to provide us with guiding principles to explore undoubtedly more complex, cognitive responses to stress. This descriptive framework represents the foundation upon which a more fully articulated theory of stress and performance can be built.

INTRODUCTION

There have been many theories concerning the relationship between stress and human performance. The longest-lived of these stems from ideas derived from the work of Yerkes and Dodson (1908). Popularity for this inverted-U description was bolstered by Hebb's (1955) influential work on drive and the conceptual nervous system. One of several fundamental problems with this traditional inverted-U concept lies in its inability to generate strong hypotheses, (see Teigen, 1994). In contrast, the 'evident' descriptive ubiquity of the inverted-U has been a strong influence on its longevity as an explanation. Eventually, the tension between descriptive strength and theoretical vacuity proved sufficiently dissonant that this simple concept lost credibility as a method of predicting stress effects (Hockey, 1983; Hancock 1987; Holland & Hancock, 1991). Following the failure of the unitary arousal account, founded on the inverted-U description, Hockey and his colleagues (see Hockey & Hamilton, 1983) provided a descriptive matrix relating the actions of various forms of environmental stress to sequential elements of human information processing. Though this approach was initially descriptive, Hockey has elaborated upon this idea in later work to provide a more complete explanation (Hockey, 1997). Partially as a result of a landmark meeting (Hockey, Gaillard, & Coles, 1986), a number of new theoretical formulations have emerged. For example, Wickens (1986) tied stress to performance through the notion of gain on information-processing stages and Hancock (1986) provided early representations of the maximal adaptability model (see also Hancock & Warm, 1989). Each of these efforts were stimulated by the fact that, at that time, there existed no satisfactory theory of stress and performance, nor even a sound model representation of this process (but see Sanders, 1983). While there have been a number of developments in the last decade (see Hancock & Desmond, 2001), here we present a prototype model which can serve as a descriptive basis for the development of a more comprehensive theory. A unique feature of the present concept is the explicit concern for both physical and cognitive forms of stress and the way in which these forms may interact in real-world situations.

THE DESCRIPTIVE MODEL

The structure of our descriptive model is shown in Figure 1. The first characterization is the input environmental demand, of which task demand is a significant component. As is evident, task demand is addressed primarily by cognitive response, and overall 'environmental' demand by physiological response (although the model also provides for ongoing cross-talk in this sequence). Physiological response to overall demand is divided into two components. First is a system-specific component in which the proximally respondent system (e.g., the thermoregulatory system for thermal effects, the cardio-respiratory system for aerobic demand, etc.) acts to meet the level of the psychological or physiological challenge. In addition, the diverse effects of any imposed stress pervade the overall physiological response system. Initial reaction to the demand by the proximal system occurs through automated response. If this first line response is sufficient, dynamic physiological stability is maintained. However, when the level of demand exceeds the response capability of the proximal system, the more resourceful general response of the overall physiological system is engaged. Successful application of this global compensation would also result in retention of dynamic stability. However, compensatory failure at this level forces the burden of physiological adaptation to the level of conscious attention and therefore necessitates some form of behavioral mitigation strategy, mediated by effortful decision-making. In general, this is composed of the stereotypical epinephrine-mediated "fight or flight" behavioral choice.

The information processing system deals with task demands in fundamentally the same manner that the physiological system deals with general environmental demands. First, information is addressed by proximal systems such as visual or auditory processing. When low-level forms of sensory and perceptual processes are unable to provide satisfactory reaction, the demand is passed through to higher processing levels. This "failure" at the specific level triggers a more encompassing appraisal of the situation, therefore this model incorporates many of the features of both the unitary and distributed versions of attentional resource models (see Kahneman, 1973; Wickens, 1980). Specifically, when collective resources are unable to process task demands using appropriate, 'automated' responses, a higher level of attentional mediation is engaged, and a more strategic form of behavioral adaptation is adopted. Behavioral adaptation can range from increased conscious effort directed to particular aspects of the task, to such coping mechanisms as load shedding within team task performance situations.

We have proposed that such load re-allocation can also be accomplished by changes in task structure and interface configuration as mediated by the technological support available through adaptive systems(see also Hancock & Chignell, 1985). Although these physiological and psychological responses have been illustrated as lock-step relations, the reality of the situation is that they are highly parallel and interactive systems. The proposed model can incorporate experimental findings of cognitive performance under physical demand (see Fatkin et al., 1990) to examine such interaction.

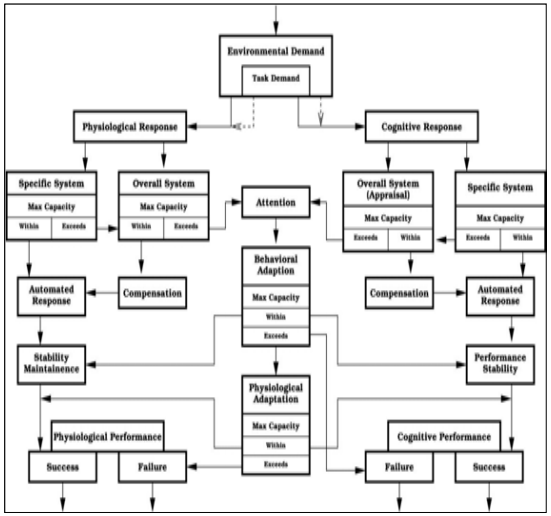


Figure 1: *A Process Model of Stress and Performance*

THE ASSESSMENT OF STRESS

Hancock and Warm (1989) have indicated that stress can be viewed as an influence on three general loci; the 'trinity of stress.' These effects can be located first as values in the environment itself, and are therefore physical measures, which are frequently used in setting limits to occupational exposure (see Hancock & Vasmatzidis, 1999). The second locus of stress effects was identified in the interaction between such environmental conditions and the individuals who are exposed to them. These represent the costs of adaptation, often referred to as 'strain effects.' Such influences can be measured directly by appraising changes in the proximal system responsible for compensatory action. Some influences, such as fluctuations in body temperature during exposure to thermal stress, are relatively easy to measure, while others, such as fatigue, are significantly more challenging due to the more complex pattern of systematic change they produce in a system. Sometimes this is easy, as in body temperature changes associated with exposure to thermal stress; sometimes it is more challenging, as in the case of fatigue, which induces a much more complex pattern of systemic change. The difficulty in the latter case is distinguishing which physiological response or collection of physiological responses best describes the condition subjectively identified as fatigue; this constitutes a serious experimental challenge. The final locus of stress effects involves the output from the stressed system or individual, and provides us with perhaps the most realistic way to measure stress effects. By measuring the change in task performance and efficiency on the primary task of concern during the application of a stressor, it is possible to directly examine the effects of the stress in real-time directly through change in performance efficiency on the primary task of concern. Sometimes this is the easiest approach, at other times, especially when future prediction is paramount, momentary changes in primary task performance are less helpful. Prediction of stress effects can be achieved using computer-based models, such as IMPRINT (Allender, Lockett, Kelley, Salvi, Mitchell, Headley, Promisel, Richer, & Feng, 1995). Therefore, each of these information sources in the 'trinity of stress' provides a different perspective from which to understand stress effects and allows specific insights into the effects of specific types of stress on psychological and physiological function.

THE ASSESSMENT OF STRESS(cont.)

In the present work, we take the vital issue of measurement one step further. We are currently engaged in a theoretical search for a common currency to describe the level of input demand. In previous work we have defined these as information rate and information structure (Hancock & Warm, 1989) which are related to the temporal and spatial characteristics of the external environment. As relational measures, they are theoretically comparable to the Gibsonian concept of 'affordances' or offerings of the environment. Thus, such measures are not simply physical values but are contingent upon functional response capacity. In the model developed in this project, we have sought to solidify this concept of 'affordance' first, in relation to known response capabilities in the face of various types and levels of stress. Thus, the majority of our experimental work will focus on the individual and combined effects of mental and physical workload. At present, we cannot predict what 'units' such measures will adopt. Certainly, they will represent ratios that will express the capability of the individual to respond versus the present level of demand. We expect to be able to use such ratio measures as inputs to both predictive modeling and subsequent mitigation strategies, such as training and adaptive systems (see Hancock & Chignell, 1987).

SUMMARY

This paper presents a descriptive framework for research upon the effects of stress on operator performance. Using this as a stepping stone, effort is made toward establishing a common metric for the various facets related to stress, as it affects operator response. Through our approach, which allows us to capture the influences of both physical and cognitive forms of stress, we hope to integrate some of the diverse representations of stress. This is achieved within the model since we posit that response processes to physiologically and psychologically challenging tasks are fundamentally identical, with respect to their mode of operation. This descriptive framework represents the foundation upon which a more fully articulated theory of stress and performance can be established.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Army Research Organization (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove and Dr. Mike Drillings for providing administrative and technical direction for the grant.

REFERENCES

(Please see handout for full list of references.)

Allender, L., Lockett, J., Kelley, T.D., Salvi, L., Mitchell, D., Headley, D.B., Promisel, D., Richer, C., & Feng, T. (1995). Verification, validation, and accreditation of a soldier-system modeling tool. *Proceedings of the Human Factors and Ergonomics Society*, **39**, 1219-1223.

Fatkin, L.T., King, J.M., & Hudgens, G.A. (1990). Evaluation of stress experienced by Yellowstone Army firefighters (Tech. Mem. No. 9-90). U.S. Army Research Laboratory, Human Research and Engineering Directorate. Aberdeen Proving Ground, MD.

Hancock, P.A., & Desmond, P.A. (Eds.). (2001). *Stress, workload and fatigue*. Lawrence Erlbaum, Mahwah: NJ.

Hancock, P.A. (1987). Arousal theory, stress and performance: Problems of incorporating energetic aspects of behavior into human-machine systems function. In: I.S. Mark, J.S. Warm, & R.L. Huston (Eds.). *Ergonomics and Human Factors: Recent Research*. (pp. 170-179). New York: Springer-Verlag.

Hockey, G.R.J. & Hamilton, P. (1983) The cognitive patterning of stress states. In: G.R.J. Hockey (Ed.). *Stress and fatigue in human performance*. New York: Wiley.

Kahneman, D. (1973). *Attention and effort* Prentice-Hall, Englewood Cliffs, NJ.

Naatanen, R. (1973). The inverted-U relationship between activation and performance: a critical review. In: S. Kornblum (Ed.). Attention and performance IV. New York: Academic Press.

Sanders, A. F. (1983). Towards a model of stress and human performance. Acta Psychologica, 53, 61-97

Weiner, E.L., Curry, R.E. & Faustina, M.L. (1984) Vigilance and taskload: In search of the inverted-U. Human Factors, 26, 215-222.

Wickens, C.D. (1986). Gain and energetics in information processing. In G.R.J. Hockey, A.W.K. Gaillard & M.G.H. Coles (Eds.). *Energetic aspects of human information processing*, Nijhoff: the Netherlands.

Yerkes, R.M. & Dodson, J.D. (1908) The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, **18**, 459-482.

Hancock, P.A., & Vasmatzidis, I. (1999). On the behavioral basis for stress exposure limits: The foundational case of thermal stress. In: W. Karwowski and W. Marras, (Eds.). *The Occupational Ergonomics Handbook*. (pp. 1707-1739), CRC Press: Boca Raton, FL.

A DESCRIPTIVE FRAMEWORK FOR THE EVALUATION OF STRESS EFFECTS ON OPERATOR PERFORMANCE

P.A. Hancock, H.C.N. Ganey, M. Mouloua, E. Salas, R. Gilson, A. Greenwood-Ericksen
University of Central Florida

R. Parasuraman
Catholic University of America

W. Harris, A. Leon
University of Minnesota

and

K. Smith
Kansas State University

ABSTRACT

In this paper we provide a general descriptive framework that relates the action of stress on operator performance capacity. The key advantage of our approach is that it allows us to capture simultaneously the influences of both physical and cognitive forms of stress and their singular and interactive effects on response efficiency. This is accomplished within the model since we propose that response processes to physiological challenge and psychological challenge are fundamentally identical in their mode of operation. By considering the commonalities of brain function with the response processes of other organs of the body, we can now use the extensive existing body of physiological insight to provide us with guiding principles to explore undoubtedly more complex cognitive responses to stress. This descriptive framework represents the foundation upon which a fully articulated theory of stress and performance is being erected.

INTRODUCTION

There have been many theories concerning the relationship between stress and human performance. The longest-lived of these stems from ideas derived from the work of Yerkes and Dodson (1908). Popularity for this inverted-U description was bolstered by Hebb's (1955) influential work on drive and the conceptual nervous system. One of several fundamental problems with this traditional inverted-U concept lies in its inability to generate strong hypotheses, (see Teigen, 1994). In contrast, the 'evident' descriptive ubiquity of the inverted-U has been a strong influence on the longevity of the arousal explanation that has been thought to underpin this relationship. Eventually, the tension between descriptive strength and theoretical vacuity proved sufficiently dissonant that this concept lost credibility as a method of predicting stress effects (Hockey, 1983; Hancock 1987; Holland & Hancock, 1991). Following the failure of the unitary arousal account, founded on the inverted-U description, Hockey and his colleagues (see Hockey & Hamilton, 1983) provided a descriptive matrix relating the actions of various forms of environmental stress to sequential elements of the human information processing sequence. Though this approach was initially descriptive, Hockey has elaborated upon this idea in later work to provide a more complete explanation (Hockey, 1997). Partially as a result of a landmark meeting (Hockey, Gaillard, & Coles, 1986), a number of new theoretical formulations have emerged. For example, Wickens (1986) tied stress to performance through the notion of gain on information-processing stages and Hancock (1986) provided early representations of the maximal adaptability model (see also Hancock & Warm, 1989). Each of these efforts were stimulated by the fact that, at that time, there existed no satisfactory theory of stress and performance, nor even a sound model representation of this process (but see Sanders, 1983). While there have been a number of developments in the last decade (see Hancock & Desmond, 2001), here we present a prototype model which can serve as a descriptive basis for the development of a more comprehensive theory. A unique feature of the present concept is the explicit concern for both physical and cognitive forms of stress and the way in which these forms may interact in real-world situations.

THE DESCRIPTIVE MODEL

The structure of our descriptive model is shown in Figure 1. The first characterization is the input environmental demand, of which task demand is a significant component. As is evident, task demand is addressed primarily by cognitive response, and overall 'environmental' demand by physiological response (although the model also provides for ongoing cross-talk in this sequence). Physiological response to overall demand is divided into two components. First is a system specific component in which the proximally respondent system (e.g., the thermoregulatory system for thermal effects, the cardio-respiratory system for aerobic demand, etc.) acts to meet the level of the physiological challenge imposed. In addition to these specific effects, the diverse effects of any stress pervade the overall physiological response system. Initial reaction to the demand by the specific or proximal system occurs through automated response. If this first line response is sufficient, dynamic physiological stability is maintained and no further compensatory response is initiated. However, when the level of demand exceeds the response capability of this proximal system, the more resourceful general response of the overall physiological system is actively engaged.

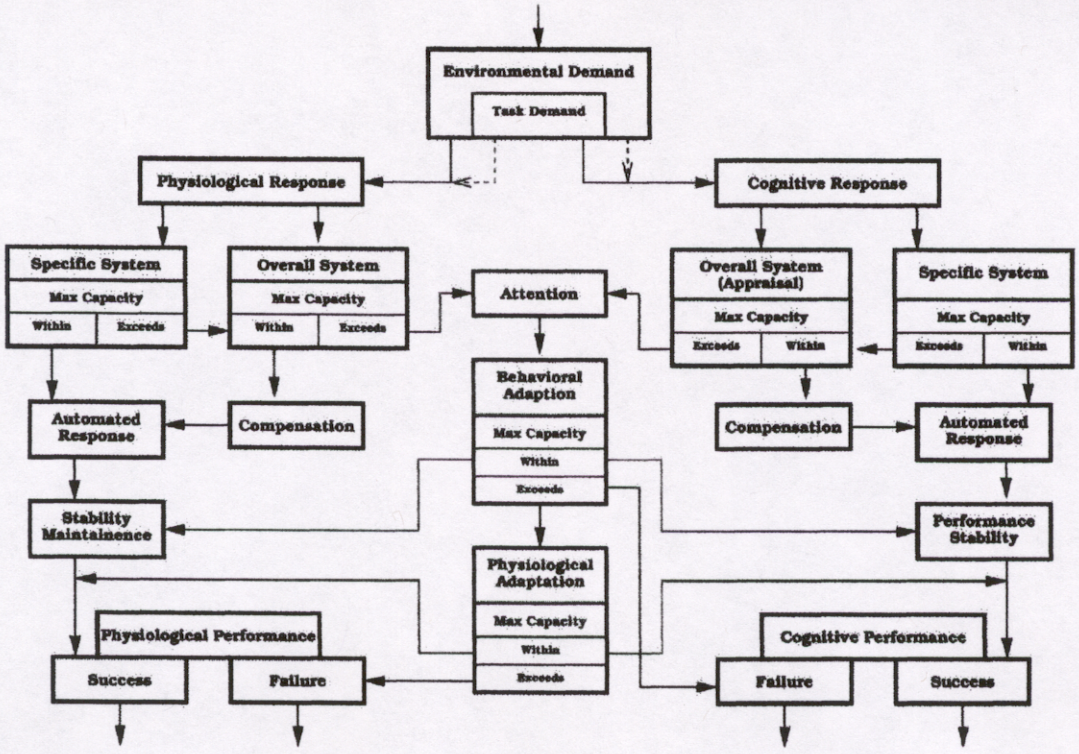


Figure 1: *A Process Model of Stress and Performance*

Successful application of this global compensation would also result in retention of dynamic stability. However, compensatory failure at this level forces the burden of physiological adaptation to the level of conscious attention and therefore necessitates some form of behavioral mitigation strategy, mediated by effortful decision-making. In general, this is composed of the stereotypical epinephrine-mediated "fight or flight" behavioral choice.

The information processing system deals with task demands in fundamentally the same manner that the physiological system deals with general environmental demands. First, information is addressed by specific proximal systems such as the visual or auditory processing systems. When low-level forms of sensory and perceptual processes are unable to provide satisfactory reaction, the demand is passed through to higher processing levels. This "failure" at the specific level triggers a more encompassing appraisal of the situation, therefore this model incorporates many of the features of both the unitary and distributed versions of attentional resource models (see Kahneman, 1973; Wickens, 1984). Specifically, when collective resources are unable to process task demands using appropriate, 'automated' responses, a higher level of attentional mediation is engaged, and a more strategic form of behavioral adaptation is adopted. Behavioral adaptation can range from increased conscious effort directed to particular aspects of the task, all the way to such coping mechanisms as load shedding within team task performance situations. We have proposed that such load re-allocation can also be accomplished by changes in task structure and interface configuration as mediated by the

technological support available through adaptive systems (see for example Hancock & Chignell, 1987). Although these physiological and psychological responses have been illustrated as lock-step relations, the reality of the situation is that they are highly parallel and interactive systems. The proposed model can thus incorporate experimental findings of cognitive performance under physical demand (see Fatkin, King, & Hudgens, 1990; Vercruyssen, Hancock, & Mihaly, 1989).

THE ASSESSMENT OF STRESS

The present descriptive conception is founded upon the earlier observations provided by Hancock and Warm (1989). They indicated that stress has three general loci of effects. These are first as values of the environment itself. These expressions are therefore commonly expressed as physical measures, which are frequently used for example in setting limits to occupational exposure (see Hancock & Vasmatzidis, 1999). The second locus of stress effects lies in the interaction between such environmental conditions and the individuals who are exposed to them. These expressions of stress represent the costs of adaptation, often referred to as 'strain effects.' Such influences can be measured directly by appraising initial changes in the proximal system responsible for compensatory action. Some influences, such as fluctuations in body temperature during exposure to thermal stress, are relatively easy to assess, while others, such as fatigue, are significantly more challenging due to the more complex pattern of systemic change they produce. The difficulty in the latter case is distinguishing which physiological response or collection of physiological responses best describes the condition subjectively identified as fatigue; this constitutes a serious experimental challenge. The final locus of stress effects involves the output from the stressed system or individual, and provides us with perhaps the most realistic and pragmatic way to evaluate stress effects. By measuring the change in task performance efficiency on the primary task of concern it is possible to directly examine the effects of that stressor in real-time. Sometimes this is the easiest approach, at other times, especially when future prediction is paramount, momentary changes in primary task performance are less helpful. Prediction of stress effects can be achieved using sophisticated computer-based models, such as IMPRINT (see Allender, Lockett, Kelley, Salvi, Mitchell, Headley, Promisel, Richer, & Feng, 1995). Therefore, each of these information sources in the 'trinity of stress' provides a different perspective from which to understand stress effects and allows specific insights into the effects of specific types of stress on psychological and physiological functioning.

Here, we take the vital issue of measurement one step further. When an individual is exposed to stress in the real world, it is very rarely contingent upon a single form of disturbance. More likely, the individual will be asked to perform a taxing sustained attention task in a hot and noisy environment, or will be asked to take critical decisions while significantly fatigued. Indeed, until we recognize the task itself as a major source of stress, we cannot complete the picture of the input that faces the exposed individual. Having done so, however, we are still faced with a seemingly intractable challenge. How do we equate the severity of these various sources. In essence, how many degrees centigrade or decibels is equivalent to a 20% increase in vigilance event rate, or additional decision alternatives? It would seem as though we are always destined to compare apples and oranges. In the present work, we are engaged in the search for a common currency to describe the level of input demand. In previous work approaches toward such a currency have attempted to distill their spatial and temporal informational basis that have thus been labeled information rate and information structure (Hancock & Warm, 1989). As relational measures, each of these notions theoretically comparable to the Gibsonian concept of an 'affordance.' As a consequence, such measures are not solely physical values but are contingent upon functional response capacity of the individual so exposed.. At present, it is uncertain what 'units' such measures will adopt. Certainly, they will represent ratios that will express the capability of the individual to respond versus the present level of demand and thus reside as measures, more closely allied to the second locus of stress. We expect to be able to use such ratio measures as inputs to both predictive modeling and subsequent mitigation strategies, such as training and adaptive systems (see Hancock & Chignell, 1987). One postulate which is proving helpful concerns the relation between stress and response variability. In the classic inverted-U function and in the later extended-U function of Hancock and Warm (1989), the description provides a relationship between the level of stress and the mean response capacity. However, it is our proposition that the variability of response also changes as a function of stress level. We here suggest that at the extremes of both high and low input demand, stress, response becomes more prototypical. That is, variability decreases as maladaptive stress increases. This observation, if supported by further empirical confirmation, provides an avenue to address the interactive stress assessment problem.

SUMMARY

This paper presents a descriptive framework for research upon the effects of stress on operator performance. Using this as a basis for development, we move to the problem of synthesizing the impact of very disparate sources of stress. Our search for a common metric to describe input sources of stress is informed by the reduction in response variability as stress increases away from an optimal region of functioning. Our general approach, which allows us to capture the influences of both physical and cognitive forms of stress, is predicated upon our observation that response processes to physiological and psychological challenge are fundamentally identical, with respect to their mode of operation. This descriptive framework represents the foundation upon which a more fully articulated theory of stress and performance may be established.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Army Research Organization (Grant# DAAD19-01-1-0621). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors wish to thank Dr. Sherry Tove and Dr. Mike Drillings for providing administrative and technical direction for the grant.

REFERENCES

Allender, L., Lockett, J., Kelley, T.D., Salvi, L., Mitchell, D., Headley, D.B., Promisel, D., Richer, C., & Feng, T. (1995). Verification, validation, and accreditation of a soldier-system modeling tool. *Proceedings of the Human Factors and Ergonomics Society*, **39**, 1219-1223.

Fatkin, L.T., King, J.M., & Hudgens, G.A. (1990). Evaluation of stress experienced by Yellowstone Army firefighters (Tech. Mem. No. 9-90). U.S. Army Research Laboratory, Human Research and Engineering Directorate. Aberdeen Proving Ground, MD.

Hancock, P.A. (1986) Stress and adaptability. In G.R.J. Hockey, A.W.K. Galliard & M.G.H.Coles (Eds.). *Energetic aspects of human information processing*, Nijhoff: the Netherlands.

Hancock, P.A. (1987). Arousal theory, stress and performance: Problems of incorporating energetic aspects of behavior into human-machine systems function. In: L.S. Mark, J.S. Warm, & R.L. Huston (Eds.). *Ergonomics and Human Factors: Recent Research.* (pp. 170-179), New York: Springer-Verlag.

Hancock, P.A. & Chignell, M.H. (1985) The principle of maximal adaptability in setting stress tolerance standards. In: R. Eberts & C. Eberts (Eds.). *Trends in Ergonomics/Human Factors II*, North Holland: Amsterdam.

Hancock, P.A., & Chignell, M.H. (1987). Adaptive control in human-machine systems. In: P.A. Hancock (Ed.). *Human Factors Psychology.* (pp. 305-345), North-Holland: Amsterdam.

Hancock, P.A., & Desmond, P.A. (Eds.). (2001). *Stress, workload and fatigue.* Lawrence Erlbaum, Mahwah: NJ.

Hancock, P.A., & Vasmatzidis, I. (1999). On the behavioral basis for stress exposure limits: The foundational case of thermal stress. In: W. Karwowski and W. Marras, (Eds.). *The Occupational Ergonomics Handbook.* (pp. 1707-1739), CRC Press: Boca Raton, FL.

Hebb, D.O. (1955) Drives and the C.N.S. (Conceptual Nervous System). *Psychological Review*, **62**, 243-254.

Hockey, G.R.J. (1983). *Stress and fatigue in human performance.* New York: Wiley

Hockey, G.R.J. (1997) Compensatory control in the regulation of human performance under stress and high workload: A cognitive- energetical framework. *Biological Psychology* **45**(1-3), p 73-93

Hockey, G.R.J. & Hamilton, P. (1983) The cognitive patterning of stress states. In: G.R.J. Hockey (Ed.). *Stress and fatigue in human performance.* New York: Wiley.

Hockey, G.R.J. Galliard, A.W.K. & Coles, M.G.H. (Eds.). (1986) *Energetic aspects of human information processing*, Nijhoff: the Netherlands.

Holland, F.G., & Hancock, P.A. The inverted-U: A paradigm in chaos. Paper presented at the Annual Meeting of NASPSPA, The North American Society for the Psychology of Sport and Physical Activity, Asilomar, CA, June, 1991.

Kahneman, D. (1973). *Attention and effort.* Englewood-Cliffs, NJ: prenstice-Hall.

Naatanen, R. (1973). The inverted-U relationship between activation and performance: a critical review. In: S. Kornblum (Ed.). *Attention and performance IV.* New York: Academic Press.

Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, **53**, 61-97

Teigen, K.H. (1994). Yerkes-Dodson: A law for all seasons. *Theory and Psychology*, **4**(4), 525-547.

Vercruyssen, M., Hancock, P.A., & Mihaly, T. (1989). Time estimation performance before, during, and following physical activity. *Journal of Human Ergology*, **18**, 169-179.

Weiner, E.L., Curry, R.E. & Faustina, M.L. (1984) Vigilance and taskload: In search of the inverted-U. *Human Factors*, **26**, 215-222.

Wickens, C.D. (1984). Processing resource in attention. In: R. Parasuraman and D.R. Davies (Eds.). *Varieties of attention.* (pp. 63-102), Orlando, FL: Academic Press.

Wickens, C.D. (1986). Gain and energetics in information processing. In G.R.J. Hockey, A.W.K. Galliard & M.G.H. Coles (Eds.). *Energetic aspects of human information processing*, Nijhoff: the Netherlands.

Yerkes, R.M. & Dodson, J.D. (1908) The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, **18**, 459-482.

THEOR. ISSUES IN ERGON. SCI., 2002, VOL. 3, NO. 2, 115–123

Ergonomics as a foundation for a science of purpose

P. A. HANCOCK* and D. D. DIAZ

Department of Psychology, 411 Phillips Hall, University of Central Florida, Orlando, FL 32816, USA

Keywords: Ergonomics; Science of purpose; Human performance.

The argument here is at once simple and profound. The first premise is that technology is the most powerful 'shaping' force on the planet today. The second premise is that individual impact on this force is most evident at the human–machine interface. A corollary of the second premise is that unity in intention is to be found in the aggregate of individual actions. Ergonomics, as the pursuit that mediates between operator and technology, is, therefore, the first step along the path to a science of intention as expressed in 'world' changes. Some initial observations on the integration of existing behavioural theories represent the first step along this royal road.

1. Introduction

For ergonomics to become more than an 'appliance science' which solely attaches heuristic patches to flawed initial designs, it has to generate and validate a utilitarian theory of human–technology interaction. With very rare exceptions, such as the conception of symvatology (Karwowski 2000), there are essentially no theories of ergonomics (see Meister 1999). Given that there promises always to be more individuals generating unhealthy technology than there are ergonomists to cure it, theory is our only viable method to power ahead of the development curve. However, what form should such a theory take and what differentiates it from any other theory of contextual behaviour? We need not be apologetic that we have no present answer, since ergonomic theorizing is in its infant stages. After all, despite one or two notable historical exceptions (Jastrzębowski 1857), ours is a small and still growing branch of knowledge and, here, we do not erect a full theory of intention as mediated by ergonomics. However, we do advance some ideas that might be used as foundation-stones for the base of such an edifice.

2. Theories in ergonomics

As evidenced by the Journal you are holding in your hand, there is a growing interest in the role of theory in ergonomics. Many scientists have recently commented either indirectly or explicitly on the unsatisfactory situation in respect to ergonomic theory (Hendrick 2000, Rasmussen 2000). In the past, ergonomics has traditionally 'borrowed' or assumed theories from foundational disciplines such as physiology and psychology (Meister 2000, Jaksholt 2000), while extrapolating methods from engineering and mathematics in order to implement the ideas and notions that have emerged from the inter-disciplinary vigour associated with the linkage between

* Author for correspondence. e-mail: phancock@pegasus.cc.ucf.edu

Theoretical Issues in Ergonomics Science ISSN 1463–922X print/ISSN 1464–536X online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI 10.1080/1463922021012379 8

humans and technology. Now, we are engaged in a search for our own theoretical basis, contingent upon but separated from the constructs of contributory disciplines. One clear instance of this search is evident in the idea of human-centred design. It is through this concept that we have sought to derive quantitative descriptions for the 'goodness of fit' between any one individual and the proximal technology of their use. Karwowski's notion of 'ergonomic entropy' is central to such a conception and we applaud his insight. However, the designation of entropic state specifies a disparity between *what is*, and *what could be*. We believe that the supraordinate and differentiating issue is the difference between *what is* and *what should be*. This denotes an explicit need for an explanation of purpose. As we have previously noted (Hancock 1997), human-centred design and, thus, 'ergonomic entropy' can well apply whether the purpose of the technology itself is suicide, genocide or terracide. Thus, for example, implementation of principles in human-centred design can obviously help terrorists achieve their aim of covert attack. It is in this manner that Moray (1993) and others have observed that science itself is a neutral pursuit with respect to the aims toward which it is directed. In contrast to this neutral posture, our primary aim here is to advance the contention that purpose itself is the critical central component of any fundamental ergonomic theory. Indeed it is, in its very essence, this concern with ethics and purpose that sets ergonomic theory apart from all other theories of behaviour in context (Hancock 1998). That this places a higher burden upon us in ergonomics to seek and specify what are ethical directions for technological purpose is a philosophical challenge, but one which elevates us beyond our foundational disciplines and helps codify our uniqueness. For, as Moray (1993: 34) has observed:

> The ultimate purpose of our discipline is not only to improve productivity or the efficiency of systems used for killing people but, rather, to better the human condition. The time has come to seriously consider what role human factors can play in the coming century. The problems are global ones, and if human factors does not play its role, we will be guilty of the *trahaison des clercs* (treason by the intellectuals) as it applies to our discipline. We have a moral duty to find our role.

We are not prepared, as yet, nor arguably is anyone else, to present a fully developed theory which features intention and purpose. We readily acknowledge that we are as yet only in the early formative stages of any such encompassing perspective. Thus, the present work is simply a first step in which we seek to show linkages between extant behavioural theories which have traditionally underpinned ergonomics. Founded on this beginning, we endeavour to show what the outlines of a theory of ergonomics purpose may be. By illuminating this one bridge between two behavioural theories, the present paper aims to serve as a catalyst for further theoretical discourse and development.

3. Purpose predicates process

In his 1857 treatise 'The Science of Work', Jastrzębowski related the potential societal, spiritual, and technological benefits attained through striving to create environments where humans can work 'at the least expenditure of toil'. Subsequent to his derivation of the name 'ergonomics', Jastrzębowski delineated many concerns fundamental to ergonomics which remain applicable today. Of especial interest is that, while Jastrzębowski outlined what a science of ergonomics *can* do, he made a specific point of describing what a science of ergonomics *should* do. He

addressed the dichotomy that would be brought by the development of a science of ergonomics: that harmful acts are facilitated to the same or to a greater degree than are beneficial acts. As evident in definitional data presented by Dempsey *et al.* (2000), the notion of purpose has all but been lost in subsequent definitions of ergonomics since Jastrzębowski's seminal work.

Having recognized the supposed neutrality of intellectual pursuits, why should we not let ethics be that field principally directed toward the designation of right and wrong? Our answer lies in the central character of ergonomics in bridging from ideas, goals, conceptions, and aspirations to their associated consequences in environmental and behavioural change. Ethics serves to study moral judgements and is concerned with *ex post facto* generation of statements about prior behaviours. Ethics as a voice is predominantly passive. In contrast, ergonomics selects and facilitates behaviours and, thereby, represents an active endeavour. While the ethicist can shed light on moralistic decisions, the ergonomist is actively capable of shaping the outcome of future behaviour. However, before aspiring to a world as we hope it could be, it is important to understand the world as it is.

4. The practical problem: ergonomics amongst rapid technological development

In our contemporary society, change is the only constant. Were this all, we might be able to cope, but unfortunately its derivative, the rate of such change, is itself in flux—a situation that renders crystallized knowledge concerning specific technical systems rapidly obsolete. Who today, other than an antique collector, needs someone to repair an eight-track tape-recorder or generate punched cards for data analysis? As software innovations replace hardware capacities, individuals become 'deskilled' even more rapidly and, in capitalist societies, such job instability generates significant social stress. In competitive economies, there is then constant focus on the new. The customer, to consume, must be persuaded that what is on offer is better than that which they currently possess. While this is often not the case, vast amounts of resources are spent on advertising to persuade enough individuals of the exact opposite, so they are induced to perceive that they now *need* what is on offer—this being an intentional manipulation of behaviour. Those who cannot fill megabytes are sold gigabytes, and suburban drivers possess vehicles that are able to go 150 mph—but never do. In this cavalcade of novelty, manufacturers have now discovered how important human–machine interfaces are and have adjusted their marketing strategies accordingly. Workstations are touted as 'ergonomic' and epithets such as 'fahrvegnugen' surface periodically to reinforce this putative advantage. None the less, especially in the field of high technology, it remains too often the case that the financial anxiety over product release overwhelms the concern for user facility. As the number of professional ergonomists is eclipsed by the exponential growth of these new technologies, it is clear that *post hoc* strategies are not a plausible means to implement the ergonomic ethos.

5. Toward an integrative theory

If collective ergonomic fingers cannot plug the flooding dyke of technological dreck, how can we ensure that any ergonomic integrity is injected into new products? As we note, purely *post hoc* means of regulation such as litigation are clearly not feasible change agents when considering the scale and the dynamics of the problem. Thus, we have to engage a more viable solution in an *a priori* approach where theory guides the design process. A caveat to this proposition is that, in the realm of the ergonomic

sciences, no theory currently exists with a wide enough breadth to account for sufficient environmental circumstances, capability requirements, design variations, and user characteristics that distinguish one innovation from another (but see Karwowski 2000). The first step toward a resolution of this problem is to coalesce fundamental contemporary theories relevant to ergonomic science in order to attain the basis for a more universally applicable theory. Theoretical convergences have been and continue to be sought (e.g. Chignell *et al.* 1999, Hancock 1996, and see also Wilson 1998). Examples can be found in general theories of interactivity of organisms with their environment and in answers to enduring questions regarding the mind–brain relationship, as well as the conceptualization of consciousness (e.g. Gomes 1995). Ergonomics would benefit greatly from an acceptable general theory; and steps toward this gain can perhaps be achieved by first reconciling two dominant approaches already used in the field of behaviour, the ecological perspective and the information processing perspective.

6. Comprehensive perspectives

Ecological psychology, at its heart, deals with the relationship between the properties of an organism and the properties of its environment. It is often viewed in stark contrast with the Cartesian vision, which separates the activities of the mind from the external world. Gibson's ecological account of perception focused primarily on vision and presented a broad theoretical framework in which response is motivated and constrained by features of the surrounding environment—for a comprehensive description, see Gibson (1954, 1966, 1979) and Greeno (1994). From a Human Factors standpoint, the ecological approach is in accord with the basic tenet of compatibility, in which minimal cognitive effort is necessarily demanded in order to successfully interact with a system (see Flach *et al.* 1995, Hancock *et al.* 1995). In contrast, the information-processing approach proposes that stimulation and cues present in the environment enter the cognitive process via the senses for subsequent transformation and processing. Thus, the ecological view focuses on what information is present in the environment, while information processing focuses on how internal representations of the environment are constructed from elementary sense data (Rosch 1996). While Gibson's theory remains phenomenological in nature and difficult to test (Blake 1994), many information-processing theories seek corroboration with neurophysiological evidence, in order to more completely explain the actual cognitive activities that mediate perception and response (see Marr 1982). Aside from their levels of specificity, the duality of mind and environment is the primary source of conflict between these two theories. Rather than representing irrevocable differences, the disparity between these two approaches may be reconcilable, so to suggest a more robust foundation upon which to erect a purpose-specific ergonomic theory.

7. Bridging existing theories

The first step to be taken in bridging any two apparently disparate theories is to search for and identify any commonalties (Hancock and Parasuraman 2002). In the present case, we take one concept derived from the very earliest foundation of ecological theory and show how, in its essence, it is synonymous with one of the most recent concepts in the energetic facet of information processing. The concept from ecological psychology is the notion of the 'field of safe travel' derived from one

of Gibson's earliest published works. The energetic concept is situation awareness (SA) (Endsley 1995).

Even before his crucial work for the US military during World War II, and well before his later articulations of ecological theory (e.g. Gibson 1966, 1979), Gibson examined the real-world problem of automobile control. In typical Gibsonian fashion, he went to the very heart of the problem and, together with his colleague Crooks, produced what still remains today one of the most influential works on driving ever published (Gibson and Crooks 1938). Clearly influenced by Lewin's *Topological Psychology* (1936), Gibson and Crooks strived to articulate two fundamental concepts that dictated vehicle control. The first, the 'minimum stopping zone' represented a purely physical value derived from the kinematics and kinetics of the vehicle involved. Given sufficient knowledge of the physical environment, the minimum stopping zone of any vehicle can always be specified. However, Gibson and Crooks also articulated the notion of the 'Field of Safe Travel', and it is this conception that dominates our attention here. It is best to explain this concept in precisely the terms given by Gibson and Crooks (1938: 454). They indicated:

> Within the boundaries of the road lies, according to our hypothesis, an indefinitely bounded field which we shall name the field of safe travel. It consists, at any given moment, of the field of possible paths which the car may take unimpeded. Phenomenally, it is a sort of tongue protruding forward along the road. Its boundaries are chiefly determined by objects or features of the terrain with a negative 'valence' in perception.

This concept, that specifies a relation between the capabilities of driver and vehicle (a human–machine system) and the opportunities intrinsic to the environment, is clearly a forerunner of the later notion of affordance that represents the central pillar of the (Gibsonian) ecological position (see Nakayama 1994, Treffner 1999). The fact that, in this example, the minimum stopping zone represents the constraints on the driver–vehicle dyad, rather than the constraints on the motor response system of the unaided individual alone is immaterial to the argument. For, as Gibson and Crooks noted, driving, like flying, skating, wind-surfing etc, is only a form of aided locomotion. Thus, the 'field of safe travel' (FST) is a concept that applies in virtually any dynamic motion situation, whether the carriage of concern is that of a single human limb or a multi-passenger airliner. Our present argument uses the concept in these general terms rather than its specific driving case. We consider the general case of the 'field of safe travel' as synonymous with 'situation awareness,' a provocative but supportable assertion.

Situation awareness is a conception that has received much recent attention and has been the subject of considerable dispute (see Flach 1995, Smith and Hancock 1995). The leading figure and proponent of situation awareness is Endsley, and it is thus fair to define the concept in her terms. She indicates that:

> Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future (Endsley 1995: 36, after Endsley 1987, 1988).

Situation awareness (SA) has been a pragmatically useful construct and has followed up on a number of comparable notions such as cognitive workload, in trying to capture the energetic state of the individual. Fortunately, there have been numerous efforts to tie the notion of situation awareness to traditional

information processing components, and those efforts are likely to loom large in importance if our theoretical bridge is solid.

First and foremost, both SA and FST are identified with perceptions and interpretations from these perceptions. Both definitions present the external environment as a source where information is contained and also exerting two types of influence. The first is that ambient environments provide distinct characteristics which stand out from the rest of the surround to motivate response and generate intent; this is denoted as 'objects or features' in FST and as 'elements' in SA. Secondly, it provides a more general level of information about the actor's surrounding in terms of space and time, which is characterized in FST as 'an indefinitely bounded fielded . . . at any given moment' and as 'a volume of time and space' in the definition of SA. The two definitions are also consistent in their idea that an actor responds to the environment based upon the consequences that are expected to follow. This notion is manifest in FST as 'a negative valance' and as 'comprehension of [the] meaning [of environmental elements]' in SA. This indicates that a response to a specific feature present in the environment occurs due to both the actor's own intent drawn from hedonistic decision-making, past experience, and also due to intent conveyed through the environment for which certain types of features intrinsically enact appetitive or repelling influence. Lastly, both definitions describe anticipative reasoning or strategizing; this is indicated by the idea in FST of use of the term 'protruding forward' and by Endsley's notion of SA as applicable to 'the near future'. This notion is driven by influences from the environment and through reasoning on the part of the actor, as well as the synergy from when the parts of the actor are compatible with parts of the environment.

Thus, when we unpack each of the critical components of the definitions of FST and SA, we can see that they are essentially co-incident and this must set critiques of SA in a different light (see Flach 1995). Further, since, after these comparisons, no additional critical terms are left in either definition, these concepts, in their general expression, are essentially synonymous. If the precursor to Gibson's affordances is synonymous with SA, some might say this is merely re-discovery. However, we believe it represents something much more important than this. It is our contention that this link represents a first step along a road to a comprehensive theory of ergonomics and intention.

8. A science of intention

Looking carefully at the definitions of each concept, we can see that they each seek to specify answers to the questions who, what, when, and where. By inference, they each address the how question through reference to any proximal technology under consideration. What each definition defaults on is 'why'. It is evident that the implied answer to the why question is linked to some general statement of intent such as the desire to get from origin to destination, but for what purpose? If the purpose is antithetical to the general good, should ergonomics support such intent? What is the compatibility between individual and collective intent using convivial and/or monopolistic tools (Illich 1973)? In essence, since purpose predicates process, process independent of purpose is mindless and, currently, our world is dominated by mindless technology. Our appeal is for a mindful technology and an integrated theory of behaviour that brings intent to the forefront. Before asking who, what, when, where, or how, the ergonomist must be the scientist who asks 'why?'

Some have analogized our relationship to burgeoning technology as a novice rides on an out-of-control mustang. If ergonomics is simply a handmaid of rampant consumerism, our greatest aspiration is akin to making sure the spurs are shiny and sharp. We are asking ergonomists not merely to evaluate the interface between man and horse but to raise its sights to consider the nature and purpose of the journey. At present, we are so taken up by simply hanging on, there are few to peruse where we are going and an even smaller number to ask why we are going in that direction. Since we believe ergonomics is the nascent form of science of intention—asking why is our primary purpose.

Acknowledgements

This research was facilitated by two grants from the United States Army. The first was the Multiple University Research Initiative—Operator Performance Under Stress (MURI-OPUS) Grant (#DAAD19-01-1-0621). The second was to the Advanced Decision Architecture Consortium (#DAAD19-01-2-0009). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. The authors would like to thank Dr Sherry Tove and Dr Mike Drillings for providing administration and technical direction for the first grant and to Drs Edwardo Salas, Jennifer Fowlkes and David Woods for the second grant. We are especially grateful to John Flach, Mica Endsley and two anonymous reviewers for their comments and suggestions with respect to the present work. While accepting many suggestions and criticisms, the wake continues to grow and develop as indeed the observations imply that it must.

References

BLAKE, R. 1994, Gibson's inspired but latent prelude to visual motion perception, *Psychological Review*, **101**, 324–328.

CHIGNELL, M. H., HANCOCK, P. A. and TAKESHITA, H. 1999, Human computer interaction: the psychology of augmented human behavior. In P. A. Hancock (ed.), *Human Performance and Ergonomics* (San Diego: Academic Press), 291–328.

DEMPSEY, P. G., WOGALTER, M. S. and HANCOCK, P. A. 2000, What's in a name? Using terms from definitions to examine the fundamental foundation of human factors and ergonomics science, *Theoretical Issues in Ergonomics Science*, **1**, 3–10.

ENDSLEY, M. R. 1987, SAGAT: a Methodology for the Measurement of Situational Awareness (NOR DOC 87–83) (Hawthrone, CA: Northrop Corp.).

ENDSLEY, M. R. 1988, Situation awareness global assessment technique (SAGAT), *Proceedings of the National Aerospace and Electronics Conference* (*NAECON*), (New York: IEEE), 789–795.

ENDSLEY, M. R. 1995, Toward a theory of situational awareness in dynamic systems, *Human Factors*, **37**, 65–84.

FLACH, J. M. 1995, Situation awareness: proceed with caution, *Human Factors*, **37**, 149–157.

FLACH, J., HANCOCK, P. A., CAIRD, J. K. and VICENTE, K. (eds) 1995, *Global Perspectives on the Ecology of Human–Machine Systems* (Mawah, NJ: Lawrence Erlbaum).

GIBSON, J. J. 1954, The visual perception of object motion and subjective movement, *Psychological Review*, **61**, 304–314.

GIBSON, J. J. 1966, *The Senses Considered as Perceptual Systems* (Boston: Houghton-Mifflin).

GIBSON, J. J. 1979, *The Ecological Approach to Visual Perception* (Boston: Houghton-Mifflin).

GIBSON, J. J. and CROOKS, L. E. 1938, A theoretical field analysis of automobile driving, *American Journal of Psychology*, **51**, 453–471.

GOMES, G. 1995, Self-awareness and the mind–body problem, *Philosophical Psychology*, **8**, 155–165.

GREENO, J. G. 1994, Gibson's affordances, *Psychological Review*, **101**, 336–342.
HANCOCK, P. A. 1996, On convergent technological evolution, *Ergonomics in Design*, **4**, 22–29.
HANCOCK, P. A. 1997, *Essays on the Future of Human–Machine Systems* (Eden Prairie, MN: Banta).
HANCOCK, P. A. 1998, Should human factors prevent or impede access? *Ergonomics in Design*, **6**, 4.
HANCOCK, P. A. and PARASURAMAN, R. 2002, Human factors and ergonomics, *Handbook of Cognitive Science*, in press.
HANCOCK, P. A., FLACH, J., CAIRD, J. K. and VICENTE, K. (eds) 1995, *Local Applications in the Ecology of Human–Machine Systems* (Mawah, NJ: Lawrence Erlbaum).
HENDRICK, H. W. 2000, The technology of ergonomics, *Theoretical Issues in Ergonomic Science*, **1**, 22–33.
ILLICH, I. 1973, *Tools for Conviviality* (New York: Harper & Row).
JAKSHOLT, T. 2000, A note on ergonomics of theory construction and use, *Theoretical Issues in Ergonomic Science*, **1**, 62–75.
JASTRZEBOWSKI, W. 1857, *An Outline of Ergonomics or Science of Work Based Upon Truths Drawn from the Science of Nature* (in Polish), [Translated into English 1997] (Warsaw: The Institute of Labour Protection).
KARWOWSKI, W., 2000, Symvatology: the science of an artifact–human compatibility, *Theoretical Issues in Ergonomics Science*, **1**, 76–91.
LEWIN, K. 1936, *Principles of Topological Psychology* [Trans. by F. Heider and G. H. Heider], (New York: McGraw-Hill).
MARR, D. 1982, *Vision* (New York: W.H. Freeman).
MEISTER, D. 1999, *The History of Human Factors and Ergonomics* (Mahwah, NJ: Erlbaum).
MEISTER, D. 2000, Theoretical issues in general and developmental ergonomics, *Theoretical Issues in Ergonomic Science*, **1**, 13–21.
MORAY, N. 1993, Technosophy and humane factors: a personal view, *Ergonomics in Design*, **1**, 33–39.
NAKAYAMA, K. 1994, James J. Gibson—an appreciation, *Psychological Review*, **101**, 329–335.
RASMUSSEN, J. 2000, Human factors in a dynamic information society: where are we heading? *Ergonomics*, **43**, 869–879.
ROSCH, E. 1996, The environment of minds: toward a noetic and hedonic ecology. In M. P. Freidman and E. C. Carterette (eds), *Cognitive Ecology* (San Diego: Academic Press), 3–27.
SMITH, K. and HANCOCK, P. A. 1995, Situation awareness is adaptive, externally-directed consciousness, *Human Factors*, **37**, 137–148.
TREFFNER, P. J. 1999, The common structure is the affordances in the ecology, *Behavioral and Brain Sciences*, **22**, 731–732.
WILSON, E. O. 1998, *Consilience: the Unity of Knowledge* (New York: Knopf).

About the authors

Peter Hancock is Provost Distinguished Research Professor in the Department of Psychology and at the Institute for Simulation and Training at the University of Central Florida. In his previous appointment, he founded and was the Director of the Human Factors Research Laboratory at the University of Minnesota. At Minnesota, he held appointments as Full Professor in the Departments of Computer Science and Electrical Engineering, Mechanical Engineering, Psychology, and Kinesiology as well as at the Cognitive Science Center and the Center on Aging Research. He currently holds a courtesy appointment as a Research Scientist at the Massachusetts Institute of Technology (MIT) and as an Adjunct Senior Research Scientist at the Transportation Institute of the University of Michigan. Professor Hancock is the author of over 200 refereed scientific articles and publications, as well as editing numerous books including: *Human Performance and Ergonomics* in the Handbook of Perception and Cognition series (Academic Press, 1999) and *Stress, Workload, and Fatigue* (Lawrence Erlbaum, 2001). He is the author of the 1997 book, *Essays on the Future of Human–Machine Systems*. He has been continuously funded by extramural sources for every year of

his professional career, including support from NASA, NIH, NIA, FAA, FHWA, the US Navy and the US Army as well as numerous State and Industrial agencies. He is the Principal Investigator on the recently awarded Multi-Disciplinary University Research Initiative, in which he will oversee $5 million of funded research on stress, workload, and performance. In 1999, he was the *Arnold Small Lecturer* of the Human Factors and Ergonomics Society and in 2000 he was awarded the *Sir Frederic Bartlett Medal* of the Ergonomics Society of Great Britain for lifetime achievement. He was the Keynote Speaker for the International Ergonomics Association and the Human Factors and Ergonomics Society at the 2000 combined meeting in San Diego. In 2001, he won the *Franklin V. Taylor Award* of the American Psychological Association as well as the *Liberty Mutual Prize for Occupational Safety and Ergonomics* from the International Ergonomics Association. In association with his colleagues Raja Parasuraman and Anthony Masalonis, he was the winner of the *Jerome Hirsch Ely Award* of the Human Factors and Ergonomics Society for 2001, the same year in which he was elected a *Fellow* of the International Ergonomics Association. His current experimental work concerns the evaluation of behavioural response to high-stress conditions. His theoretical works concerns human relations with technology and the possible futures of this symbiosis. He is a *Fellow* of and past President of the Human Factors and Ergonomics Society. He collects and studies antique maps and is a committed Ricardian.

Derek Diaz is a Doctoral Candidate in Psychology at the University of Central Florida. His prime interests are in Human Factors/Ergonomics and the linkage with wider aspects of human behaviour.

Quarterly Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield

Army MURI, Subcontract to U. Central Florida

Raja Parasuraman
Catholic University of America

PERFORMANCE PERIOD: 12/1/01- 3/31/02

During this period, three studies were presented at the Midyear meeting of the American Psychological Association, Divisions 19 and 21, "Contemporary & Emerging Issues in Human Factors, Engineering and Military Psychology" at Ft. Belvoir, VA, March 7-8, 2002.

Papers Presented

1. Rovira, E., Zinni, M., & Parasuraman, R. *Information and decision* at *uncertainty: Effects of automation unreliability on operator performance and workload.*

This study demonstrated the adverse but differential effects of both information and decision uncertainty on operator performance. Results were interpreted within the framework of a stages of automation model of Parasuraman, Sheridan, & Wickens (2000). Future work is planned to use adaptive automation to mitigate the adverse performance effects noted in this study.

A paper based on this study has been accepted for presentation at the Annual Meeting of the Human Factors and Ergonomics Society, Baltimore, September 2002, and will appear in the proceedings of that conference.

2. Harris, W.C., Parasuraman, R., Zinni, M., Hancock, P.A., & Harris, S. C. *Viking: Development of an infantry scenario multi-task environment.*

This study described the development of the Viking task platform. Viking allows for close experimental manipulation of task factors relevant to investigating operator performance under conditions of stress, workload, and fatigue. Initial data supported the sensitivity of the task platform to factors relevant to infantry soldier performance in the electronic battlefield.

3. Rovira, E., & Parasuraman, R. *Sensor to Shooter: Task Development & Empirical Evaluation of Effects of Automation Unreliability*

This study examined the effects of four different levels of decision support automation on performance on a simulation of a "sensor-to-fire" battlefield system (STS). In a dynamic display with moving friendly, enemy and unknown units, operators had to decide whether an enemy could be fired upon effectively, while maintaining communication and conserving resources, protecting headquarters, and using artillery and battalion units appropriately. Independent variables included the type and level of decision-aiding support (complete listing, priority listing, top choices, and recommendation of decision choice) and automation reliability (60%, 80%). Dependent

variables included accuracy and reaction time of engagement decisions. Under the higher reliability condition (80%) there was a greater cost in accuracy for higher levels of automation (priority listing, top choice, and recommendation) than at a lower level (complete listing). The results supported the view that automation unreliability has a greater performance cost for decision automation than for information automation.

A paper based on this study has also been accepted for presentation at the Annual Meeting of the Human Factors and Ergonomics Society, Baltimore, September 2002, and will appear in the proceedings of that conference.

Ongoing Studies

1. The Viking platform has been modified so that automation support can be applied at two stages in the Parasuraman, Sheridan, and Wickens (2000) model of automation: information and decision support. An initial study examining the effects of sub-task difficulty, sustained work, and automation on performance is currently being conducted and will be completed over the summer.

2. A study examining the use of adaptive automation using real-time measurement of mental workload has been completed, but data analysis is ongoing. An adaptive algorithm (or adaptive logic) was empirically derived from changes in the 0.1 Hz measure of heart rate variability (HRV) and used to trigger changes in automation state in response to measured mental workload in real time during a 90-minute session performing the MAT task. The results will be examined with respect to a workload-matched model of adaptation proposed by Parasuraman, Mouloua, and Hilburn (1999; see also Parasuraman & Hancock, 2001).

(This study was also partially supported under a separate grant from NASA Langley Research Center).

References

PARASURAMAN, R., & HANCOCK, P. A. (2001). Adaptive control of workload. In P. A. Hancock & P. E. Desmond (Eds.) *Stress, Workload, and Fatigue.* (pp. 305-320). Mahwah, NJ: Erlbaum.

PARASURAMAN, R., MOULOUA, M., & HILBURN, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. W. Scerbo & M. Mouloua (Eds.) *Automation Technology and Human Performance: Current Research and Trends.* (pp. 119-123). Mahwah, NJ: Erlbaum.

PARASURAMAN, SHERIDAN, T. B., & WICKENS, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, 30*, 286-297.

Initial Progress Report

Date: 08/01/01

To: Dr. Peter Hancock, Principal Investigator, University of Central Florida

From: Drs. Arthur Leon, Principal Investigator & Ava Walker, Project Coordinator
University of Minnesota
Laboratory of Physiological Hygiene and Exercise

Re: MURI Subcontract Agreement w/University of Minnesota
under Army #DAAO –19-01-1-0621
"Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield"

1. Overview. Primary areas of responsibility for this initial phase of the project included organization and logistics tasks, literature search and review, protocol development, institutional review board application preparation

2. Organization and Logistics tasks. Every effort is being made to coordinate between Laboratory (Drs. Leon and Walker) and Field-testing (Dr. Wayne Harris) concerns and requirements.

a. administrative tasks related to contract agreement and personnel issues
b. equipment and facilities identified and secured
c. materials and supplies identified
d. computer needs assessed but not ordered at this time
e. coordination with U of MN/Fairview Medical Center Biochemistry Laboratory for salivary tests

3. Literature Search and Review.

a. extensive computer search
    1) performed electronic database search for scientific research articles, books and technical reports on the topic: physical exertion and cognitive function
    2) most pertinent articles have been acquired and are in the process of being reviewed
b. detailed summary of the current scientific knowledge is being developed regarding this topic, as well as a search for strategies to mitigate mental and physical stress

4. Protocol and Assessment Technique Development.

Based on a review of the literature, assessment techniques have been selected, and study protocols for all research activities, from recruitment to data storage are being finalized.

a. recruiting activities include the identification and initial contact of potential volunteers for the project (i.e. ROTC students and the U of MN and members of the MN Army National Guard)
b. baseline fitness assessment and cognitive function protocols are being finalized
c. VO_2max (maximal oxygen consumption) testing: selection of treadmill for graded exercise testing protocol

1) Balke Treadmill Protocol – 3.4 MPH walking with military issue back pack (Ruck Sack), with a 2% increase in % grade every 3 minutes until volitional exhaustion
2) critical variables (i.e. VO_2, VCO_2, ECG, HR, HR variability, BP, RQ, RPE)

d. dose-response treadmill exercise protocol: a one-hour "simulated road-march" testing protocol with participant walking at 3.4 mph at 75% of their VO_2 max is being developed with multiple sampling opportunities to allow the assessment of cognitive function throughout the endurance-related activity
e. assessment of basic cognitive function will use the Automated Neuropsychological Assessment Metrics (ANAM) battery.
f. salivary amylase and cortisol assays to determine physiological fatigue are being studied and protocols for the assessment of are being investigated through the U of MN/Fairview Medical Center Biochemistry Laboratory ($17/per sample)

5. Institutional Review Board/Human Subjects Application Preparation:
IRB application is being prepared during this initial phase and will be submitted NLT mid-August.

6. Projections for mid-phase I
a. completion of Literature summary report
b. IRB approval and begin pilot project with human volunteers by end of September.

Respectfully Submitted,

Arthur S. Leon, MD Ava J. Walker, Ph.D

Interim Progress Report

Date: 09/25/01

To: Dr. Peter Hancock, Principal Investigator, University of Central Florida

From: Drs. Arthur Leon, Principal Investigator & Ava Walker, Project Coordinator
University of Minnesota
Laboratory of Physiological Hygiene and Exercise

Re: MURI Subcontract Agreement w/University of Minnesota
under Army #DAAO –19-01-1-0621
"Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield"

Work moves forward here at the University of Minnesota as we wait for final approval from the full Institutional Review Board on Human Subjects Research.

- Full IRB committee meets tomorrow, Sept 26th and we are prepared to answer any and all questions quickly so as to meet with full approval NLT October 2, 2001.
- The full literature search is nearing completion both in the area of cognitive function and physiological stress, and in the area of factors that might improve cognitive function during long-term physiological stress.
- Protocols and data collection procedures are being rehearsed and refined.
- Letters of cooperation were sent out to all branches of ROTC, and National Guard and Reserve Commanders and we met with all ROTC commanders regarding involving their soldiers.
- Orientation packets and forms are ready.
- Data collection for ten pilot subjects projected to be completed by mid- November.
- Materials such as salivary vials and POMS questionnaires have been received.
- Equipment yet to be obtained is a computer in the Exercise Science Lab.

Respectfully Submitted,

Arthur S. Leon, MD Ava J. Walker, Ph.D

U of MN MURI Phase I - Progress Report_December

Date: 12/21/01

To: Dr. Peter Hancock, Principal Investigator, University of Central Florida

From: Drs. Arthur Leon, Principal Investigator & Ava Walker, Project Coordinator
University of Minnesota, Laboratory of Physiological Hygiene and Exercise

Re: MURI Subcontract Agreement w/University of Minnesota under Army #DAAO –19-01-1-0621 "Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield"

1. Overview. Primary areas of responsibility for the MN Center during Phase I included organization and logistics tasks, literature search and review, protocol and Manual of Procedure development, recruiting strategies and, institutional review board approval and a Pilot Physical Activity study involving 10 subjects. The following tasks were accomplished during Phase I:

2. Organization and Logistics tasks. Task that were completed include:
   a. administrative tasks related to contract agreement and personnel issues
   b. equipment and facilities identified and secured
   c. materials and supplies identified and secured in cooperation with Dr. Harris
   d. computer needs assessed but not ordered at this time**
   e. account set up and procedures modified with U of MN/Fairview Medical Center Biochemistry Laboratory for salivary tests
   f. IRB approval
   g. completion of data collection in a pilot study
   h. protocol development for Phase II.

3. Literature Search and Review.
   a. Extensive computer search
      1) performed electronic database search for scientific research articles, books and technical reports on the topic: physical exertion and cognitive function
      2) most pertinent articles have been acquired and are in the process of being reviewed
   b. Detailed summary of the current scientific knowledge was developed regarding the topic of physical fatigue and cognitive function, as well as a search to identify strategies to mitigate mental and physical stress

4. Protocol and Assessment Technique Development.
Based on a review of the literature, recruitment strategies, and assessment techniques were selected, and study protocols for all research activities, from to data storage are finalized.
   a. recruiting activities included the identification and initial contact of potential volunteers for the project (i.e. ROTC cadet at the U of MN, active and reserve service members on duty locally (including MN Army & Air National Guard, Army and Air Reservist)
   b. baseline fitness assessment and cognitive function protocols were finalized
   c. VO_2max (maximal oxygen consumption) testing: selection of treadmill protocol for graded exercise testing
      1) Balke Treadmill Protocol – 3.4 MPH carrying a military issue back pack

(Ruck Sack), with a 2% increase in % grade every 3 minutes until volitional exhaustion**

2) physiological variables (i.e. VO_2, VCO_2, ECG, HR, BP, RQ, RPE)

d. dose-response treadmill exercise protocol: a one-hour "simulated road-march" testing protocol with participant walking at 3.4 mph at 75% of their VO_2 max is being developed with multiple sampling opportunities to allow the assessment of cognitive function throughout the endurance-related activity

e. assessment of basic cognitive function during exercise testing will use the Automated Neuropsychological Assessment Metrics (ANAM) battery.

f. salivary amylase and cortisol assays to verify physiological fatigue performed by the U of MN/Fairview Medical Center Biochemistry Laboratory ($17/per sample)

5. Pilot Study Completed. See informed consent and protocol documents for complete details. Briefly:

a. 21 participants were successfully recruited from military personnel in the area

b. Orientation. 10 participants (3 females, 7 males) were selected to participate during the abbreviated time period in Phase I, signed informed consents, were given a complete orientation to the testing protocols and provided with detailed instruction on the ANAM test battery, monitored through on entire session and provided a practice disk with written instructions and directed to practice 3 more times before returning to the Lab.

c. Visit 1. Participants completed the POMS questionnaire, then were fitted with a 12 lead ECG, monitored to HR, ECG, blood pressure, RPE, VO2 utilization through a maximal graded exercise test walking 3.4 mph carry a 35lb ruck sack on their back. Participants began walking on the treadmill, performed a pre-test ANAM test then began the Max test protocol of a 2% increase every two minutes until volitional exhaustion. They then cooled down for 2 minutes continued to walk on the treadmill and performed a post–test ANAM. Saliva samples were collected within the last minute of each ANAM session.

d. Visits 2 & 3. Participants came back to the Lab 2 more times for exercise testing in order to explore the dose-response of cognitive function in two different intensities and durations of stressful physical activity – treadmill walking with a 35 lb ruck sack.

1) Visit 2. They performed a graded exercise test involving 3 levels of intensity (at HR associated with 30%, 60% and 90% respectively) of their VO2max values. Each time they performed the ANAM. HR, RPE, and saliva samples were obtained.

2) Visit 3. The last visit involved performing a 5-mile road march at 3.4 mph carry a 35 lb ruck sack. Participant performed a pre-post POMS, the completed an ANAM session at time 0, 20, 40, 60 & 80 minutes during the march. HR, RPE, and saliva samples were taken during the beginning and end of each ANAM session.

e. Data retrieval from the ANAM, database creation and data entry are complete, lab results from saliva sample are lagging but expected this week. Preliminary data analysis has begun and preliminary reports have been drafted.

6. Future Projects: At the start of Phase II. 1) Secure a laptop computer to be used for data collection during exercise. 2) Modify protocol to change walk pace from 3.4 mph to 3.7 mph during all visits. 3) Modify protocol to randomly assign the treatment order – Intensity and Road March. 4) Change and resubmit IRB to reflect changes in protocol. 5) Finish data analysis from the pilot study and prepare data for presentation at a regional APA conference in March and perhaps another scientific meeting.

Respectfully Submitted, Arthur S. Leon, MD Ava J. Walker, Ph.D

Mitigating Stress, Workload and Fatigue on the Electronic Battlefield

DAAD 19-01-1-0621

Initial Progress Report: 1 August 2001

Kansas State University

Professor Kip Smith

Progress had been made on two fronts since 16 June 2001 start date of the grant. First, we attended the meeting with DOD and other members of the grant team on 25 June 2001. Ms. Lauren Murphy briefed the group on the scope of our planned work and our initial experimental design. Second, we are taking steps to begin the experimentation. These steps include modifying our Human Subjects Committee approvals to reflect the plan of work presented at the June meeting and developing the computer software to display the stimuli for the extended form game. Experimentation will begin with the arrival of the student body at the end of the month.

Kansas State University
Prof. Kip Smith

Overview

We have completed a pair of studies on operator performance under varying levels of task-induced stress. We consider task demand to be a stressor (source of stress) while stress itself to be the operator's response to stressors. Task demand is often the major source of task-relevant stress for operators working in complex, dynamic environments.

1 Lauren Murphy & Kip Smith. A study of task demand, response to task demand and performance.

This study has been completed with N=6 subjects. The domain was air-traffic control. The experimental platform was a realistic simulation of an air traffic controller's workstation. The study documented the relationships between (a) task demand, (b) the operator's psychophysiologic and behavioral response to that demand, and (c) the operator performance. The study controlled the level of task demand (aircraft traffic count) in a series of realistic en-route air traffic scenarios and monitored participants' heart rates, communications, traffic-controlling actions, and number of operational errors. Operational errors were found to be equally likely after a peak in task demand as they were during a peak. Heart-rate variability was found to reflect the quality of the operator's performance and not the demands of the task environment: it varied monotonically with the occurrence of errors and deviated systematically from the pattern of traffic count. These findings suggest that, in a simulation environment, heart-rate variability is a robust indicator of the relative level of operator stress (response to task demand.)

2 Nicole Peck & Kip Smith Emotional responses to information in a multi-agent environment. (This study was also partially funded by a grant from NSF.)

This study has been completed with N=24 subjects. The task consisted of bidding with real money for goods at auction. The experimental platform was a computer simulation of (a) the 'Dutch clock' used in Aalsmeer Holland to sell flowers and bulbs to wholesalers, and (b) its inverse, the 'English clock,' used by Sotheby's and others to sell goods to the public. In English auction, prices rise (values shown on the clock increase); the last person in the auction buys the good. In the Dutch auction, prices fall (values shown on the clock decrease); the first person to jump into the auction buys the good. While both environments present a potentially stressful situation, the English is relatively static; the Dutch is more dynamic and, presumably, more stressful.

The study explored the response to task demand (stress) across a pair of potentially stressful environments, the English and Dutch auctions. Groups of four participants bid in 60 Dutch and 60 English auctions. Heart rate was used

as a physiological measure of emotional response to task demand. Results suggest that (a) the receipt of a token (task-relevant information in support of a goal) elicits an increase in heart-rate, and (b) the relinquishing of a token (achievement of a goal) elicits a decrease in heart rate. We interpret the relative levels of heart rate as a proxy for anxiety, an emotional indicator of stress. These findings suggest that, in an economic environment involving the exchange of cash, heart-rate is a robust indicator of the relative level of operator stress (response to task demand.)

In addition to these studies, we are engaged in the following task platform/simulation development efforts.

1 Kendall, Shelton & Smith. Software development to simulate the navigation of a tugboat in a narrow twisting channel. The task requires cooperation between a pilot (steering) and an engineer (velocity) to avoid crashing the boat into the river bank. The platform allows experiments to control (a) the delay in the feedback between commands and the boat's response, (b) the sinuosity of the channel, and (c) the level of reward for successfully reaching the dock. System development is complete. We plan to use this platform to measure the effects of stress on cooperative task performance.

2 Shelton & Smith. Software development to allow distributed play of the extensive form game (McCabe & Smith, 2000), a two-person economic game with a sub-game perfect equilibrium that gives Player 1 an opportunity to trust Player 2 and Player 2 a choice between reciprocity and defection. System development will be complete by the end of November. We plan to use this platform to allow participants to induce in each other differing levels of trust-relevant stress: Players 1 who trust Players 2 who then defect will likely experience more trust-induced stress than Players 1 who trust Players 2 who reciprocate.

3 Murphy, Peck & Smith. Software development to present the Metzler and Shepard mental rotation task, a well-documented test of spatial abilities. We intend to use this platform to measure the effects of stress on spatial abilities. This experiment is ready and waiting for permission to collect data.

4 McMillen & Smith. Software development to present the Hobbits and Orcs problem-solving task. We intend to use this platform to measure the effects of stress on individual and cooperative problem solving. This work is getting under way.

STANDARD

NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION

AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

AGREEMENT TITLE: Operator Under Stress Research

AGREEMENT NUMBER: NCRADA-**NAVAIR ORLANDO TSD-02-037**

AGREEMENT ADMINISTRATORS:

NAVAIR ORLANDO TSD

Technology Transfer Office:	**Ms. Patricia R. Nimmo, AIR 49T, (407) 380-8225**
Legal Counsel:	**Robert W. Adams, AIR 772P, (407) 380-8211**
Principal Investigator:	**Ron Wolff, AIR 4962, (407) 380-4583**

UCF MURI / OPUS

Preferred Contact:	**Mrs. Kim Smith, UCF Office of Research, (407) 823-3062**
Principal Investigator:	**Dr. Peter Hancock, Director, UCF MURI/OPUS Laboratory, (407) 823-2310**

This Page Intentionally Left Blank

TABLE OF CONTENTS

PREAMBLE

Article 1.
DEFINITIONS 2

1.1 "Agreement" 2
1.2 "Classified Information" 2
1.3 "Collaborator" 2
1.4 "Controlled Unclassified Information (CUI)" 2
1.5 "Cooperative Work" 2
1.6 "Data" 2
1.7 "Effective Date" 2
1.8 "Exclusive License" 2
1.9 "Government" 2
1.10 "Government Purpose Rights" 2
1.11 "Information" 2
1.12 "Intellectual Property" 2
1.13 "Invention" 2
1.14 "Invention Disclosure" 3
1.15 "Made" 3
1.16 "Militarily Critical Technologies (MCT)" 3
1.17 "Non-Subject Data" 3
1.18 "Non-Subject Invention" 3
1.19 "Patent Application" 3
1.20 "Principal Investigator (PI)" 3
1.21 "Proprietary Information" 3
1.22 "Restricted Access Information" 3
1.23 "Subject Data" 3
1.24 "Subject Invention" 3
1.25 "Tangible Property" 4
1.26 "Unlimited Rights" 4

Article 2. OBJECTIVES 4

Article 3. RESPONSIBILITIES 4

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities 4
3.2 **UCF MURI / OPUS Laboratory** Personnel and facilities 5
3.3 Security Regulations and Directives 5

Article 4. REPRESENTATIONS AND WARRANTIES 5

4.1 **NAVAIR ORLANDO TSD** Representations and Warranties 5
4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Representations and Warranties 5
4.3 Joint Representations 6

Article 5. FUNDING 7

5.1 Payment Schedule..... 7
5.2 Insufficient and Excess Funds..... 7
5.3 No New Commitments..... 8
5.4 Accounting Records..... 8

Article 6. REPORTS AND PUBLICATIONS..... 8

6.1 Final Reports..... 8
6.2 Agreement to Confer Prior to Publication or Public Disclosure..... 9
6.3 Classified Information 9

Article 7. INTELLECTUAL PROPERTY..... 9

7.1 Data..... 9
7.1.1 General Provisions Applying to All Data..... 9
7.1.1.1 Ownership..... 9
7.1.1.2 No Implied License..... 9
7.1.1.3 Marking of Data..... 10
7.1.1.3.1 Data Provided With Less Than Unlimited Rights.. 10
7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information..... 10
7.1.1.3.3 Data That are Subject to 35 USC 205..... 11
7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted..... 11
7.1.1.4 Protection of Data..... 11
7.1.1.5 Release of Data Under the Freedom of Information Act..... 11
7.1.2 Subject Data..... 12
7.1.2.1 Delivery of Requested Subject Data..... 12
7.1.2.2 Rights in Subject Data..... 12
7.1.3 Rights in Non-Subject Data..... 12
7.2 Copyrights..... 13
7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory**..... 13
7.2.2 Copyright License to the Government..... 13
7.2.3 Copyright Statement..... 13
7.3 Trademarks and Service Marks..... 13
7.3.1 Ownership of Trademarks and Service Marks..... 13
7.3.2 Obligation of Employees to Report Trademarks and Service Marks.... 13
7.3.3 Obligation of Collaborators to Notify Each Other..... 13
7.3.4 Responsibility for Filing an Application for Trademark or Service Mark..... 14
7.3.5 License to Use Trademark or Service Mark..... 14
7.4 Subject Inventions..... 14
7.4.1 Obligation to Report Subject Inventions..... 14
7.4.1.1 Collaborators' Instruction to Employees..... 14
7.4.1.2 Timely Invention Disclosure by Inventors..... 14
7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator..... 14
7.4.2 Determination of Subject Inventions..... 14
7.4.3 Title to and Ownership of Subject Inventions..... 14

7.4.4 Filing of Patent Applications..... 15
7.4.4.1 Filing of Patent Applications on Solely Made Inventions..... 15
7.4.4.2 Filing of Patent Applications on Jointly Made Inventions..... 15
7.4.4.3 Preserving Intellectual Property Rights..... 15
7.4.4.4 Filing Deadlines..... 15
7.4.4.5 Copies and Inspection..... 15
7.4.4.5.1 Copies of Prosecution Papers..... 15
7.4.4.5.2 Access to Patent Application File and Right to Make Copies..... 16
7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application..... 16
7.4.5 Nonexclusive License to Subject Inventions..... 16
7.4.5.1 Nonexclusive License Grant..... 16
7.4.5.2 Confirmatory Nonexclusive License Agreement..... 16
7.4.6 Option for Exclusive License to Subject Inventions..... 16
7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement. 17
7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions..... 17
7.4.8.1 Exclusive Licenses and Exclusive License Option..... 17
7.4.8.2 Nonexclusive Licenses..... 17
7.5 Non-Subject Inventions..... 18
7.5.1 Ownership of Non-Subject Inventions..... 18
7.5.2 Rights Under Other Agreements..... 18
7.5.3 No License to Non-Subject Inventions..... 18
7.6 Research License..... 18

Article 8. TANGIBLE PROPERTY..... 18

8.1 Title to Preexisting Tangible Property..... 18
8.2 Tangible Property Purchased by Collaborators to Perform Cooperative Work... 18
8.3 Title to Developed Tangible Property..... 19
8.4 Tangible Property Operational and Disposition Costs..... 19
8.5 Disposal of Tangible Property..... 19

Article 9. LIABILITY..... 19

9.1 Extent of Government Liability..... 19
9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Liability..... 19
9.3 *Force Majeure*..... 20

Article 10. GENERAL PROVISIONS..... 20

10.1 Characteristics of the Agreement..... 20
10.1.1 Entire Agreement..... 20
10.1.2 Severability..... 20
10.1.3 Interpretation of Headings..... 20
10.2 Agreement Between Collaborators..... 20
10.2.1 Governing Laws..... 20
10.2.2 Independent Parties/Entities..... 20
10.2.3 Assignment /Subcontracting..... 20

10.2.4 Disputes..... 21
10.2.4.1Settlement and Resolution..... 21
10.2.4.2Continuation of Cooperative Work..... 21
10.2.5 Waivers..... 21
10.2.6 Use of Name or Endorsements..... 22
10.3 Environment, Safety, and Health..... 22
10.4 U.S. Competitiveness..... 22
10.5 Public Release of This Agreement..... 22

Article 11. MODIFICATIONS AND NOTICES..... 22

11.1 Amendments..... 22
11.2 Termination..... 23
11.2.1 Termination by Mutual Consent..... 23
11.2.2 Unilateral Termination..... 23
11.3 Notices..... 23

Article 12. SURVIVING PROVISIONS..... 24

Article 13. DURATION..... 24

Article 14. SIGNATURES..... 24

Appendix A
Statement of Work

Appendix B
Confirmatory License Agreement

STANDARD
NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

PREAMBLE

Under authority of the U.S. Federal Technology Transfer Act of 1986 (Public Law 99-502, 20 October 1986, as amended), **NAVAIR ORLANDO TSD (Naval Air Warfare Center Training Systems Division)**, located at **12350 Research Parkway, Orlando, Florida 32826-3275**, and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, whose administrative offices are located at **12443 Research Parkway, Suite 207, Orlando, FL 32826-3252**, enter into this Cooperative Research and Development Agreement (CRADA), which shall be binding upon the Collaborators and their assignees according to the clauses and conditions hereof and for the term and duration set forth.

The U.S. Federal Technology Transfer Act of 1986, as amended, provides for making the expertise, capabilities, and technologies of U.S. Federal laboratories accessible to other Federal agencies; units of State or local government; industrial organizations (including corporations, partnerships and limited partnerships, and industrial development organizations); public and private foundations; nonprofit organizations (including universities); or other persons in order to improve the economic, environmental, and social well-being of the United States by stimulating utilization of U.S. Federally funded technology developments and/or capabilities.

NAVAIR ORLANDO TSD has extensive expertise, capabilities, and information in **Virtual Environments for Operator Training**, and, in accordance with the U.S. Federal Technology Transfer Act, desires to make this expertise and technology available for use in the public and private sectors.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory has the interest, resources, capabilities, and technical expertise to transition the results of Naval research and development for public use.

NOW THEREFORE, the Collaborators agree as follows.

Article 1. DEFINITIONS

As used in this Agreement, the following terms shall have the meanings defined below, which are equally applicable to both the singular and plural forms of nouns or any tense of verbs.

1.1 "Agreement" means this Cooperative Research and Development Agreement (CRADA) with its Appendices.

1.2 "Classified Information" means all Data classified in accordance with the national security laws of the United States.

1.3 "Collaborator" means the Navy participant or the Non-Navy participant represented and bound by the signatories of this Agreement.

1.4 "Controlled Unclassified Information (CUI)" means Government Data, Information, or materials provided to or resulting from this Agreement that may be export controlled, sensitive, for official use only, or otherwise protected by law, executive order, or regulation.

1.5 "Cooperative Work" means research, development, engineering, or other tasks performed under this Agreement by **NAVAIR ORLANDO TSD** or **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** working individually or together, pursuant to the Objectives (Article 2) and the Statement of Work (Appendix A).

1.6 "Data" means recorded information of any kind regardless of the form or method of the recording, including computer software.

1.7 "Effective Date" means the date of the last signature of the Collaborators executing this Agreement.

1.8 "Exclusive License" means the grant by the owner of Intellectual Property of the exclusive right to make, use, or sell a patented invention.

1.9 "Government" means the Government of the United States of America.

1.10 "Government Purpose Rights" means the right of the Government to use, duplicate, or disclose Data, in whole or in part, and in any manner, for Government purposes only, and to have or permit others to do so for Government purposes only. Government Purpose Rights includes competitive procurement, but does not include the right to have or permit others to use Data for commercial purposes.

1.11 "Information" means all data, trade secrets, and commercial and financial information. (Chapter 5 Subsection II of Title 5 USC)

1.12 "Intellectual Property" means the property of ideas, examples of which include, but are not limited to, patents, trademarks, copyrights, and trade secrets.

1.13 "Invention" means any invention or discovery that is or may be patentable or otherwise protected under Title 35, United States Code, or any novel variety of plant that is or may be patentable under the Plant Variety Protection Act. (15 USC 3703(9))

1.14 "Invention Disclosure" means the document identifying and describing to organizational management the Making of an Invention.

1.15 "Made" when used in conjunction with any Invention means the conception or first actual reduction to practice of such Invention. (15 USC 3703(10))

1.16 "Militarily Critical Technologies (MCT)" means those technologies identified in the Militarily Critical Technologies List and under the Export Administration Act of 1979, as amended.

1.17 "Non-Subject Data" means any Data that are not Subject Data.

1.18 "Non-Subject Invention" means any Invention that is not a Subject Invention.

1.19 "Patent Application" means an application for patent protection for an Invention with any domestic or foreign patent-issuing authority.

1.20 "Principal Investigator (PI)" means that person having the responsibility for the performance of the Cooperative Work on behalf of a Collaborator.

1.21 "Proprietary Information" means information that embodies trade secrets developed at private expense or business, commercial, or financial information that is privileged or confidential provided that such information:

is not known or available from other sources without obligations concerning its confidentiality;

has not been made available by the owners to others without obligation concerning its confidentiality;

is not already available to the Government without obligation concerning its confidentiality; and

has not been developed independently by persons who have had no access to the information. (FAR/DFARS Definition)

1.22 "Restricted Access Information" means Subject Data generated by **NAVAIR ORLANDO TSD** that would be Proprietary Information if the Information had been obtained from a non-Federal Collaborator participating in a CRADA (15 USC 3710a). Under 15 USC 3710a(c)(7)(B), the Collaborators mutually may agree to provide appropriate protection to Subject Data generated by **NAVAIR ORLANDO TSD** (Restricted Access Information) against public dissemination or release under the Freedom of Information Act (FOIA) for a period of up to five (5) years after development of the Information.

1.23 "Subject Data" means that Data first recorded in the performance of the Cooperative Work.

1.24 "Subject Invention" means any Invention Made in the performance of the Cooperative Work.

1.25 "Tangible Property" means personal or real property that can be physically touched or held.

1.26 "Unlimited Rights" means the right to use, modify, reproduce, release, disclose, perform, or display Data or Computer Programs in whole or in part, in any manner and for any purpose whatsoever, and to have or permit others to do so.

Article 2. OBJECTIVES

Background

NAVAIR ORLANDO TSD has been conducting Navy sponsored research in virtual environments for operator/small arms training for over 15 years. Results of this research have already been transitioned for many military training applications, and have also been the basis for technology transition in use by law enforcement. The Small Arms Simulation Training (SAST) laboratory at **NAVAIR ORLANDO TSD** now has in use an extensive range of software and data bases that have resulted from in-house **NAVAIR ORLANDO TSD** research work.

The University of Central Florida (UCF), with plans to establish a new laboratory for Human Factors research related to Homeland Defense, became aware of the work already done in **NAVAIR ORLANDO TSD's** SAST laboratory. Discussions were immediately begun to explore the mutual benefits of an agreement to work collaboratively.

The primary objective of this work will be to enhance mutual understand of operator performance under stress and develop possible applications of human factors research in this area for military training, and for homeland defense and related technologies

Benefit to Navy Partner: Research using **NAVAIR ORLANDO TSD** developed or enhanced source code may be expected to result in further enhancements to software and databases, at no expense to the government, and with direct use in military simulation-based training. Additionally, the anticipated UCF human factors research projects may be expected to have direct application and benefit for military training. Further, expertise will be provided to UCF on a cost basis, leveraging Navy research dollars.

Benefit to Non-Navy partner: **NAVAIR ORLANDO TSD** developed or enhanced source code for scenario generation, rural and urban databases, weapon tracking, image generation, networking/communication, etc., can provide the kind of virtual environment test bed required for planned human factors research, and can greatly accelerate work toward research objectives.

Article 3. RESPONSIBILITIES

The Collaborators shall provide personnel, facilities, and equipment necessary for, and shall perform, the Cooperative Work.

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities

The Cooperative Work done by **NAVAIR ORLANDO TSD** will be performed under the program guidance of **Ron Wolff, PI, NAVAIR ORLANDO TSD, AIR 4962**, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed

within the facilities of **NAVAIR ORLANDO TSD** or done on behalf of **NAVAIR ORLANDO TSD** by third parties in support of this Agreement.

UCF MURI / OPUS Laboratory personnel who perform Cooperative Work at **NAVAIR ORLANDO TSD** facilities will be supervised by the **UCF MURI / OPUS Laboratory** PI.

3.2 **UCF MURI / OPUS Laboratory** Personnel and Facilities

The Cooperative Work done by the **UCF MURI / OPUS Laboratory** will be performed under the program guidance of **Dr. Peter Hancock, UCF MURI / OPUS Laboratory** PI, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed within the facilities of the **UCF MURI / OPUS Laboratory** or done on behalf of the **UCF MURI / OPUS Laboratory** by third parties in support of this Agreement.

NAVAIR ORLANDO TSD personnel who perform Cooperative Work at the **UCF MURI / OPUS Laboratory** facilities will be supervised by **NAVAIR ORLANDO TSD** PI.

3.3 Security Regulations and Directives

Each Collaborator will abide by the safety and security regulations and directives of the host facility in which the Cooperative Work is being performed.

Article 4. REPRESENTATIONS AND WARRANTIES

4.1 **NAVAIR ORLANDO TSD**'s Representations and Warranties

NAVAIR ORLANDO TSD hereby warrants and represents to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** as follows:

4.1.1 **NAVAIR ORLANDO TSD** is a Federal laboratory of the U.S. Department of the Navy (Navy) as defined by 15 USC 3710a(d)(2)(A) and Department of Defense Instruction 5535.8, dated May 14, 1999.

4.1.2 The performance of the activities specified by this Agreement is consistent with the **Virtual Environments for Operator/Small Arms Training** and technology transfer missions of **NAVAIR ORLANDO TSD** (15 USC 3710a).

4.1.3 The Department of the Navy official executing this Agreement for **NAVAIR ORLANDO TSD** has the requisite power and authority to enter into this Agreement and to bind **NAVAIR ORLANDO TSD** to perform according to the terms of this Agreement.

4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Representations and Warranties

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory hereby warrants and represents to **NAVAIR ORLANDO TSD** as follows:

4.2.1 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not directly or indirectly controlled by a foreign company or government (Executive Order 12591, Section 4 (a)). **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, as of the Effective Date of this Agreement, is a part of a **university** duly organized, validly existing, and in good standing under the laws of **the State of Florida.**

4.2.2 The official executing this Agreement for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** has the requisite power and authority to enter into this Agreement and to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to perform according to the terms of this Agreement.

4.2.3 The University of Central Florida and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** have taken all actions required by law, or otherwise, to authorize the execution and delivery of agreements, such as this Agreement.

4.2.4 The execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any agreement binding on **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. Furthermore, the execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any valid order of any court, or any regulatory agency or other body having authority to which **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is subject.

4.2.5 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not presently subject to debarment or suspension by any agency of the Government. Should **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** be debarred or suspended during the term of this Agreement or thereafter, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** will notify **NAVAIR ORLANDO TSD** within thirty (30) days of receipt of a final notice. **NAVAIR ORLANDO TSD** may then elect to terminate this Agreement and any licenses and options granted under this Agreement.

4.2.6 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is not** a small business as defined in 15 USC 632 and implementing regulations (13 CFR 121.101 et seq.) of the Administrator of the Small Business Administration.

4.3 Joint Representations

The Collaborators make the following representations.

4.3.1 There is no express or implied warranty as to any research, Invention, or product, whether tangible or intangible. In particular, the Collaborators make no express or implied warranty as to the merchantability or fitness for a particular purpose of any research, Invention, or product, whether tangible or intangible. Likewise, the Collaborators make no express or implied warranty as to any Cooperative Work, Subject Invention, Subject Data, or other product resulting from the Cooperative Work.

4.3.2 The use and dissemination of Information and materials exchanged under this Agreement will be in accordance with all U.S. laws and regulations, including those pertaining to national security and export control. Nothing in this Agreement shall be construed as a license to export Information or to permit any disclosure in violation of law, regulation, or Department of Defense policy. The exporting Collaborator is responsible for obtaining any export licenses that may be required by U.S. Federal law.

Article 5. FUNDING

5.1 Payment Schedule

Consistent with the information in Appendix A, **NAVAIR ORLANDO TSD** will provide source code/databases developed or enhanced by **NAVAIR ORLANDO TSD** to the **MURI / OPUS** lab at no cost to **UCF**. If, at a later date during the period of this CRADA, there is a mutually determined need for additional expertise from **NAVAIR ORLANDO TSD**, then the following will apply. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees to pay **NAVAIR ORLANDO TSD** on a cost basis, as mutually determined to be required, during the period of this CRADA, as specified in Appendix A.

Checks will be payable to:

NAVAIR ORLANDO TSD, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments

Each check and its cover correspondence shall refer to Navy CRADA number "NCRADA-**NAVAIR ORLANDO TSD-02-037**."

Checks will be mailed to:

NAVAIR ORLANDO TSD, 12350 Research Parkway (AIR 4.9T Nimmo), Orlando, Florida 32826-3275, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments.

5.2 Insufficient and Excess Funds

NAVAIR ORLANDO TSD may discontinue performance under this Agreement if the funds provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** for performance by **NAVAIR ORLANDO TSD** are insufficient or are not provided as specified in Article 5.1. In the event **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to tender the Government the required payment within fifteen (15) days after its respective due date, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall be in default under this Agreement for failure to make payments. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on**

behalf of the MURI / OPUS Laboratory is in default for this reason, NAVAIR ORLANDO TSD shall notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** does not cure the default within fifteen (15) days of mailing date of notice, **NAVAIR ORLANDO TSD** may proceed to terminate the Agreement in accordance with Article 11.2.2, may cancel any option for an Exclusive License to a Subject Invention, and may terminate any Exclusive License granted pursuant to this Agreement.

Funds that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** paid under Article 5.1 and that **NAVAIR ORLANDO TSD** has not obligated or expended at the time of completion, expiration, or termination of this Agreement shall be returned to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** after **NAVAIR ORLANDO TSD's** submission of a final fiscal report to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**.

5.3 No New Commitments

NAVAIR ORLANDO TSD shall make no new commitments concerning this Agreement after receipt of a written termination notice from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in accordance with Article 11.2 and shall, to the extent practicable, cancel all outstanding commitments by the termination date. Should such cancellation result in any costs incurred by **NAVAIR ORLANDO TSD, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that such costs shall be chargeable against any funding that it provided to **NAVAIR ORLANDO TSD.**

5.4 Accounting Records

NAVAIR ORLANDO TSD shall maintain current accounts, records, and other evidence supporting all its expenditures against funding provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement and shall retain such records for at least twelve (12) months after the completion, expiration, or termination of this Agreement. **NAVAIR ORLANDO TSD** shall provide **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** a financial report within four (4) months after completion, expiration, or termination of this Agreement.

Article 6. REPORTS AND PUBLICATIONS

6.1 Final Report

The Collaborators shall submit to each other a final report within four (4) months of the completion, termination, or expiration of this Agreement that includes the results obtained and a list of all Subject Inventions Made.

6.2 Agreement to Confer Prior to Publication or Public Disclosure

The Collaborators agree to confer and consult prior to any publication or public disclosure of Subject Data to ensure that no Proprietary Information, Restricted Access Information, Government Classified Information, CUI, or MCT Information is released and that patent rights are not compromised. Prior to any such publication or public disclosure of Subject Data, each Collaborator shall be offered a period not less than fifteen (15) days and not to exceed thirty (30) days, unless otherwise mutually agreed in writing by the Collaborators, to review any proposed abstract, publication, presentation, or other document for public disclosure that contains Subject Data. For the purposes of this Article, the term "disclosure" shall include, but not be limited to, submission of any manuscript for peer review prior to publication. It is the responsibility of the Collaborator intending to make public disclosure of Subject Data to notify the other Collaborator of such intent.

If a Collaborator objects to a proposed public disclosure, that Collaborator must so notify the other Collaborator within thirty (30) days of the date of notice of intent to disclose publicly. If no objection is received by the Collaborator intending to make public disclosure, concurrence is assumed. If a Collaborator objects on the grounds that patent rights may be compromised, a Patent Application must be filed by the responsible Collaborator within ninety (90) days of the date of notification of intent to make public disclosure, or by another date mutually agreed to by the Collaborators. If a Collaborator objects to the release of Information on the grounds that the Information is Proprietary Information, Restricted Access Information, or Information whose dissemination is restricted by U.S. security laws or regulations, the disclosure shall be postponed until the Information no longer meets the definitions of Proprietary Information, Restricted Access Information, or is no longer covered by U.S. security laws or regulations.

6.3 Classified Information

Any presentation that includes Subject Data that are Classified Information or otherwise restricted Data must have prior review and approval by **NAVAIR ORLANDO TSD** pursuant to the pertinent security laws, regulations, and directives.

Article 7. INTELLECTUAL PROPERTY

7.1 Data

7.1.1 General Provisions Applying to All Data

7.1.1.1 Ownership

Each Collaborator shall have title to all Data generated by that Collaborator.

7.1.1.2 No Implied License

Unless otherwise specifically provided, the Collaborators agree that the exchange of Data of any kind does not confer a license to any Invention claimed in any

patent or Patent Application or to the subject matter of any copyright, trademark/service mark, or other form of Intellectual Property protection.

7.1.1.3 Marking of Data

7.1.1.3.1 Data Provided With Less Than Unlimited Rights

Each Collaborator shall mark all Data that it provides with less than Unlimited Rights with a marking that clearly identifies the limited rights.

7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall place a proper proprietary marking on each medium used for recording Data that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** delivers to **NAVAIR ORLANDO TSD** under this Agreement that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** asserts is Proprietary Information. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall request in writing if it wishes Subject Data generated by **NAVAIR ORLANDO TSD** to be marked as Restricted Access Information. The Collaborators together shall confer to determine if such marking is appropriate, with reference to the Definitions of Article 1. If the Collaborators mutually agree to the marking then:

(a) For Non-Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory – NAVAIR ORLANDO TSD** MAY USE ONLY FOR PURPOSE OF CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(b) For Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** – GOVERNMENT HAS GOVERNMENT PURPOSE RIGHTS UNDER CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(c) For Data that are Restricted Access Information, the marking shall read:

"RESTRICTED ACCESS INFORMATION – PROTECT IN ACCORDANCE WITH CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**" UNTIL **SEPTEMBER 2005**".

7.1.1.3.3 Data That are Subject to 35 USC 205

NAVAIR ORLANDO TSD shall mark Data it provides under this Agreement that disclose one or more Inventions in which the Government owns or may own a right, title or interest, and that are subject to confidentiality under 35 USC 205. Such Data shall be marked:

"**NAVAIR ORLANDO TSD** DATA PROTECTED FROM RELEASE OR DISCLOSURE UNDER 35 USC 205."

7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted

Each Collaborator shall mark all Data that are Classified Information, CUI, MCT, or otherwise restricted by U.S. security or export control laws or regulations that it provides under this Agreement.

7.1.1.4 Protection of Data

Except for the rights granted in Article 7.1.2.2, Data shall be protected in accordance with the proper markings of its owner and as provided by, at a minimum, the requirements of 15 USC 3710a. Proprietary Information will be protected only if it is properly marked as such. Information provided in intangible form that is Proprietary Information must be designated Proprietary Information at the time it is delivered, followed within fifteen (15) days by a writing summarizing the exact Information to be protected. The Collaborator receiving Information in an intangible form that is designated as Proprietary Information shall be responsible for protecting the Information as Proprietary Information during the fifteen (15) day notification period. After the fifteen (15) day period, if no written summary has been received, the receiving Collaborator need not continue to protect the Information received in intangible form.

Restricted Access Information shall be protected from public dissemination for up to five (5) years, as mutually agreed.

Classified Information, CUI, MCT, or otherwise restricted Information shall be protected in accordance with the security laws of the United States.

7.1.1.5 Release of Data Under the Freedom of Information Act

Data in the possession of **NAVAIR ORLANDO TSD** that are not marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or Restricted Access Information must be released by **NAVAIR ORLANDO TSD** where such release is required pursuant to a request under the Freedom of Information Act (FOIA) (5 USC 552). **NAVAIR ORLANDO TSD** shall protect Data that are properly marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI**

/ **OPUS Laboratory** or Restricted Access Information from release under the FOIA for as long as the marked Data meet the definition of CUI, Proprietary Information or Restricted Access Information. Prior to release of any such Data, **NAVAIR ORLANDO TSD** shall promptly notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** of any request for Data of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** regardless of whether the requested Data are marked Proprietary Information or as required under the public records law.

7.1.2 Subject Data

7.1.2.1 Delivery of Requested Subject Data

Each Collaborator shall have the right to review and receive delivery of all Subject Data generated by the other Collaborator. Requested Subject Data shall be delivered to the requesting Collaborator within fifteen (15) days of the request.

7.1.2.2 Rights in Subject Data

Except as represented in Article 4.3.2, the Collaborators shall have Unlimited Rights in all Subject Data that are not Proprietary Information or Restricted Access Information. Notwithstanding 15 USC 3710a, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** grants Government Purpose Rights in any Subject Data furnished by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to **NAVAIR ORLANDO TSD** under this Agreement that are properly marked as Proprietary Information. The Government has Government Purpose Rights in Subject Data that are Restricted Access Information.

7.1.3 Rights in Non-Subject Data

The Collaborators shall have Unlimited Rights in any Non-Subject Data delivered under this Agreement that are not Proprietary Information.

NAVAIR ORLANDO TSD has a limited right to use, reproduce, and disclose only to Government employees for use in support of the Cooperative Work any Non-Subject Data that are properly marked as Proprietary Information and are provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement. Such Proprietary Information can be used only for the purpose of performing the Cooperative Work unless consent to other use or disclosure is obtained from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in writing.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall have a limited right to use, reproduce, or disclose Non-Subject Data that may describe one or more Inventions in which the Government owns or may own a right, title or interest, if such Non-Subject Data are provided by **NAVAIR ORLANDO TSD** under this Agreement. In accordance with 35 USC 205, such Non-Subject Data are to be held in confidence. Such Non-Subject Data shall be properly marked by **NAVAIR ORLANDO TSD** and the limited rights of **The UCF Office of Research, UCF on behalf of its**

Board of Trustees on behalf of the MURI / OPUS Laboratory shall be defined by a separate non-disclosure agreement.

7.2 Copyrights

7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory may copyright works of authorship prepared pursuant to this Agreement if eligible for copyright protection under Title 17 USC.

7.2.2 Copyright License to the Government

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory grants to the Government a nonexclusive, irrevocable, paid-up license in copyrighted works of authorship, including software (17 USC 106) prepared pursuant to this Agreement for any purpose, consistent with the rights in Data described in Article 7.1.

7.2.3 Copyright Statement

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall include the following statement on any text, drawing, mask work or other work of authorship, that may be copyrighted under 17 USC, that is created in the performance of this Agreement:

"The U.S. Government has a copyright license in this work pursuant to a Cooperative Research and Development Agreement with **NAVAIR ORLANDO TSD.**"

7.3 Trademarks and Service Marks

7.3.1 Ownership of Trademarks and Service Marks

The Collaborator first establishing a trademark or service mark for goods or services with which the mark is used shall be considered the owner of the mark.

7.3.2 Obligation of Employees to Report Trademarks and Service Marks

Employees of both Collaborators shall report the adoption of a trademark or service mark associated with the Cooperative Work to their employer within thirty (30) days of the first use of the mark. Use includes internal use of any product or service of the Cooperative Work.

7.3.3 Obligation of Collaborators to Notify Each Other

Each Collaborator shall notify the other Collaborator within thirty (30) days of their employee's report of the first use of a trademark or service mark.

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

Each Collaborator shall be entitled to own the Subject Inventions of its employees. Each Collaborator shall cooperate with the other Collaborator to obtain inventor signatures on Patent Applications, assignments or other documents required to secure Intellectual Property protection. For any Invention Made jointly by employees of the Collaborators, each Collaborator shall have ownership of the Subject Invention in the form of an undivided interest.

7.4.4 Filing of Patent Applications

7.4.4.1 Filing of Patent Applications on Solely Made Inventions

Each Collaborator has primary responsibility for filing Patent Applications on the Subject Inventions of its employee(s).

Notwithstanding such primary responsibility, by mutual agreement, the Collaborators may identify which Collaborator shall file a Patent Application on any Subject Invention.

7.4.4.2 Filing of Patent Applications on Jointly Made Inventions

In the case of an Invention jointly Made by employees of both Collaborators, the Collaborators shall confer and agree as to which Collaborator will file any Patent Application. Officers of the non-filing Collaborator shall cooperate with the filing Collaborator to obtain signatures on documents that are needed to file a Patent Application.

7.4.4.3 Preserving Intellectual Property Rights

The Collaborator responsible for filing of a Patent Application on any Subject Invention shall file such Patent Application at least sixty (60) days prior to any bar date or one year from the date the Invention Disclosure was received, whichever comes first. If no Patent Application is filed within the specified time period, the other Collaborator may assume control of filing the Patent Application and take title to the Subject Invention on ten (10) days written notification. The Collaborator that relinquished the responsibility to file shall retain a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Subject Invention practiced throughout the world by or on its behalf.

7.4.4.4. Filing Deadlines

The Collaborator responsible for filing any Patent Application for a Subject Invention shall notify the other Collaborator of all filing deadlines for prosecution of any Patent Application and maintenance of any patents on the Subject Invention. Notwithstanding the primary responsibility defined in Article 7.4.4.1, sixty (60) days prior to any filing deadline, the Collaborators shall confer to determine if the filing Collaborator intends to respond to the filing deadline. The non-filing Collaborator will be permitted to take action if the filing Collaborator declines.

7.4.4.5 Copies and Inspection

7.4.4.5.1 Copies of Prosecution Papers

Each Collaborator filing a Patent Application on a Subject Invention shall provide the other Collaborator with a copy of any communication relating

to prosecution of said Patent Application within thirty (30) days of receipt of such communication.

7.4.4.5.2 Access to Patent Application File and Right to Make Copies

Upon written request, the filing Collaborator shall give the other Collaborator an Associate Power of Attorney, with authorization to access the Patent Application, make copies, and, in the event the filing Collaborator fails or declines to take action, do all that is necessary to secure Intellectual Property protection for the Subject Invention.

7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application

In the event both Collaborators decline to file a Patent Application on a Subject Invention, the Government will renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

In the event both Collaborators decline to file a Patent Application on a Subject Invention, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** may, at its sole discretion, renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

7.4.5 Nonexclusive License to Subject Inventions

7.4.5.1 Nonexclusive License Grant

Each Collaborator grants to the other Collaborator a nonexclusive, irrevocable, paid-up license to practice a Subject Invention Made by employees of the granting Collaborator or have the Subject Invention practiced throughout the world by or on behalf of the other Collaborator. No nonexclusive license granted under this Agreement shall permit licensee to grant sublicenses.

7.4.5.2 Confirmatory Nonexclusive License Agreement

Each Collaborator has the obligation to provide a Confirmatory License Agreement (Appendix B) to the other Collaborator for each nonexclusive license within ninety (90) days of the date of filing.

7.4.6 Option for Exclusive License to Subject Inventions

NAVAIR ORLANDO TSD gives **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** the option of acquiring an Exclusive License for the field of use **(virtual environments for operator/small arms training)** in the Government's rights in any Subject Invention Made in whole or in part by a **NAVAIR ORLANDO TSD** employee. The license shall be for reasonable consideration. In

order to exercise this option, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must notify **NAVAIR ORLANDO TSD** in writing within one hundred and eighty (180) days of the filing of a Patent Application. Unless another time period is mutually agreed upon between the Collaborators, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must execute an Exclusive License to the Subject Invention within one hundred and eighty (180) days of election to exercise the option, or the Invention shall be made available for licensing by the public in accordance with 37 CFR Part 404.

Any Exclusive License granted by the Government in a Subject Invention is subject to the statutorily required reservation by the Government of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the Government (15 USC 3710a).

7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement

No license granted under this Agreement shall be assigned, licensed or otherwise disposed of except to the successor in interest of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which such license pertains.

7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions

7.4.8.1 Exclusive Licenses and Exclusive License Option

NAVAIR ORLANDO TSD may terminate any Exclusive License or cancel any option for an Exclusive License to a Subject Invention granted under this Agreement in the event that:

(a) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is in default for failure to make payment as agreed in Article 5; or

(b) The Agreement is terminated unilaterally by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**; or

(c) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to perform according to the Statement of Work (Appendix A); or

(d) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** becomes a foreign owned, controlled, or influenced (FOCI) organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.4.8.2 Nonexclusive Licenses

NAVAIR ORLANDO TSD shall terminate any nonexclusive license to a Subject Invention granted under this Agreement if **The UCF Office of Research,**

UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory becomes a FOCI organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.5 Non-Subject Inventions

7.5.1 Ownership of Non-Subject Inventions

Each Collaborator owns its Non-Subject Inventions.

7.5.2 Rights Under Other Agreements

Nothing in this Agreement is intended to change the rights in Intellectual Property acquired by the Collaborators in any other contract or agreement between the **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** and the Government.

7.5.3 No License to Non-Subject Inventions

This Agreement does not grant any Collaborator a license, express or implied, to any Non-Subject Invention.

7.6 Research License

Each Collaborator shall allow the other Collaborator to practice any of its Non-Subject Inventions for the purpose of performing the Cooperative Work.

No license, express or implied, for commercial application(s) is granted to either Collaborator in Non-Subject Inventions by performing the Cooperative Work.

For commercial application(s) of Non-Subject Inventions, a license must be obtained from the owner.

Article 8. TANGIBLE PROPERTY

8.1 Title to Preexisting Tangible Property

Each Collaborator shall retain title to all Tangible Property to which it had title prior to the Effective Date of this Agreement.

8.2 Tangible Property Purchased by Collaborators to Perform the Cooperative Work

Each Collaborator shall retain title to all Tangible Property that it purchases during the period of this Agreement. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** cannot take title to any Government Tangible Property under this Agreement. Collaborator consumables to be used in the Cooperative Work of this Agreement are the property of the purchasing Collaborator until consumed.

8.3 Title to Developed Tangible Property

All Tangible Property developed under this Agreement with all components purchased by one Collaborator shall be the property of that Collaborator. Tangible Property having any component purchased by **NAVAIR ORLANDO TSD** shall be the property of the Government, unless such Tangible Property can reasonably be separated without damage to the other individual components. After this Agreement is completed, expired, or terminated, if separation of components can be made without damage, the Collaborators may, by mutual agreement, separate the Tangible Property into its components and the separated components shall remain the property of the Collaborator that purchased them.

8.4 Tangible Property Operational and Disposition Costs

During the period of and upon completion, expiration, or termination of this Agreement, each Collaborator shall be responsible for all costs of maintenance, removal, storage, repair, disposal, and shipping of all Tangible Property to which it has title.

8.5 Disposal of Tangible Property

Disposal of Tangible Property shall be in accordance with applicable U.S. Federal, State, and local property disposal laws, environmental laws, and regulations.

Article 9. LIABILITY

9.1 Extent of Government Liability

The Government shall be solely liable for the negligent or wrongful acts of its officers and employees to the extent provided for in the Federal Tort Claims Act (28 USC 2671 et. seq.) and in other applicable laws and regulations of the United States that specifically waive sovereign immunity. Nothing in this Agreement shall be construed as a waiver of the sovereign immunity of the United States.

9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Liability

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is solely responsible for its actions and the actions of those acting for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in the performance of this Agreement and for any damages that may arise from any suit, action, or claim, and for any costs from or incidental to any suit, action, or claim, including but not limited to settlement and defense costs. Further, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that in any suit, action or claim brought by anyone not a party to this Agreement based on actions of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not pursue any actions to enter the Government as a party in such suit, action or claim unless the Government has some liability under the Federal Tort Claims Act.

9.3 *Force Majeure*

No Collaborator shall be liable for the consequences of any *force majeure* that (1) is beyond its reasonable control; (2) is not caused by the fault or negligence of such Collaborator; (3) causes such Collaborator to be unable to perform its obligations under this Agreement; and (4) cannot be overcome by the exercise of due diligence. In the event of the occurrence of a *force majeure*, the Collaborator unable to perform shall promptly notify the other Collaborator. The Collaborators shall suspend performance only for such period of time as is necessary to overcome the result(s) of the *force majeure* and shall use their best efforts to resume performance as quickly as possible.

Article 10. GENERAL PROVISIONS

10.1 Characteristics of the Agreement

10.1.1 Entire Agreement

This Agreement constitutes the entire agreement between the Collaborators concerning the Cooperative Work and supersedes any prior understanding or written or oral agreement relative to the Cooperative Work.

10.1.2 Severability

The illegality or invalidity of any Article of this Agreement shall not impair, affect, or invalidate any other Article of this Agreement.

10.1.3 Interpretation of Headings

Headings of the Articles of this Agreement are for convenience of reference only and do not form a part of this Agreement and shall in no way affect the interpretation thereof.

10.2 Agreements Between Collaborators

10.2.1 Governing Laws

United States Federal Laws shall govern this Agreement for all purposes.

10.2.2 Independent Parties/Entities

The relationship of the Collaborators to this Agreement is that of independent parties and not as agents of each other, partners, or participants in a joint venture. Each Collaborator shall maintain sole and exclusive control over its personnel and operations.

10.2.3 Assignment/Subcontracting

10.2.3.1 Neither Collaborator may allow third parties to perform any part of the Cooperative Work under this Agreement without express written consent of the other

Collaborator. If consent is obtained, the Collaborator requesting such consent shall remain fully responsible for the portion of the Cooperative Work to be accomplished under a third-party agreement, and the third party is not a Collaborator of this Agreement. Any third-party agreement to perform a portion of the Cooperative Work shall contain terms consistent with this Agreement.

10.2.3.2 This Agreement shall not be assigned or otherwise transferred by either Collaborator without the prior written consent of the other Collaborator, except to the successor of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which this Agreement pertains.

10.2.3.3 If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee is a U.S. company, and becomes, during the term of this Agreement or thereafter, directly or indirectly owned, controlled, or influenced by a foreign company or government (FOCI), then **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee shall promptly notify **NAVAIR ORLANDO TSD** to that effect.

10.2.4 Disputes

10.2.4.1 Settlement and Resolution

NAVAIR ORLANDO TSD and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agree to use reasonable efforts to reach a fair settlement of any dispute. If such efforts are unsuccessful, remaining issues in dispute will be referred to the signatories or their successors for resolution. If a dispute continues, the remaining issues may be submitted to the Chief of Naval Research (CNR), or the CNR designee, for resolution. This Agreement does not prevent any Collaborator from pursuing disputes in a U.S. Federal court of competent jurisdiction. No Collaborator will pursue litigation in a U.S. Federal court until after the CNR, or the CNR designee, decides the dispute, or until sixty (60) days after the dispute was first submitted to the CNR, or the CNR designee, whichever comes first.

10.2.4.2 Continuation of Cooperative Work

If payments or installment payments are to be made as stated under Article 5, **NAVAIR ORLANDO TSD** will not start or continue cooperative work until payments or installment payments are received.

10.2.5 Waivers

None of the provisions of this Agreement shall be considered waived by either Collaborator unless such waiver is given in writing to the other Collaborator, signed by the executing official of this Agreement or the official's successor having the authority to bind the Collaborator making the waiver. The failure of either Collaborator to insist upon strict performance of any of the terms and conditions herein, or failure or delay to exercise any rights provided herein or by law shall not be deemed a waiver of any right of either Collaborator under this Agreement.

10.2.6 Use of Name or Endorsements

Except as provided for in Article 7.2.3, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not use the name of **NAVAIR ORLANDO TSD** or any other Government entity on any product or service that is directly or indirectly related to either this Agreement or any patent license or assignment associated with this Agreement without the prior approval of **NAVAIR ORLANDO TSD**. By entering into this Agreement, **NAVAIR ORLANDO TSD** does not directly or indirectly endorse any product or service provided, or to be provided, by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, its successors, assignees, or licensees. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not in any way imply that the Department of the Navy endorses any such product or service.

10.3 Environment, Safety, and Health

Each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste in its custody during the course of this Agreement. At the conclusion of this Agreement, each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste still in its possession. Each Collaborator shall obtain at its own expense all necessary permits and licenses as required by U.S. Federal, State, and local law and shall conduct such handling, control, and disposition in a lawful and environmentally responsible manner. Each Collaborator is responsible for all required environmental, safety, and health compliance, notice, and monitoring related to its facility in accordance with U.S. Federal, State, and local law and regulations. Collaborators shall abide by the environmental, safety, and health directives of the host facility in which the Cooperative Work is being performed, and any U.S. Federal, State, or local laws and regulations pertaining to environment, safety, and health that are applicable to the host facility.

10.4 U.S. Competitiveness

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory agrees that any product, process, or service using Intellectual Property arising from the performance of this Agreement shall be manufactured substantially in the United States.

10.5 Public Release of This Agreement

This Agreement, without funding information (Article 5) and Appendices, may be released to the public.

Article 11. MODIFICATIONS AND NOTICES

11.1 Amendments

If a Collaborator wishes to modify this Agreement, the Collaborators shall confer in good faith to determine the desirability of such modification. Such modification shall not be effective until a written amendment is signed by both executing officials of this Agreement or their successors.

11.2 Termination

11.2.1 Termination by Mutual Consent

The Collaborators may elect to terminate this Agreement at any time by mutual consent. Such termination shall not be effective until a written termination agreement is signed by both executing officials of this Agreement or their successors.

11.2.2 Unilateral Termination

A Collaborator may unilaterally terminate this entire Agreement at any time by giving the other Collaborator written notice signed by the executing official of this Agreement or his/her successor, not less than thirty (30) days prior to the desired termination date. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** unilaterally terminates this Agreement, any option for an Exclusive License to a Subject Invention and any Exclusive License to a Subject Invention granted by or pursuant to this Agreement shall simultaneously be terminated.

11.3 Notices

All notices pertaining to or required by Articles of this Agreement, except those pertaining solely to the prosecution of any patent, trademark, or service mark, shall be in writing and shall be signed by an authorized representative of the Technology Transfer Office for **NAVAIR ORLANDO TSD** or the preferred contact for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, and all such notices shall be delivered by hand, sent by courier with proper registration, or sent by certified mail, return receipt requested, with postage prepaid, addressed as follows:

If to **NAVAIR ORLANDO TSD**:

Ms. Patricia R. Nimmo
Technology Transfer Project Manager
NAVAIR ORLANDO TSD (AIR 49T)
12350 Research Parkway
Orlando, FL 32826-3275

If to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

Ms. Kim Smith
Senior Contract Specialist
Office of Research
University of Central Florida
Orlando Tech Center
12443 Research Parkway, Suite 207
Orlando, Fl 32826-3252

A Collaborator shall notify the other Collaborator of a change of address in the manner set forth above.

Notices pertaining solely to the prosecution of any patent, trademark, or service mark related to this Agreement shall be in writing and shall be signed by and sent to the Collaborator's legal counsel for Intellectual Property. Legal counsel for Intellectual Property for each Collaborator shall send a copy of any such notice to the Technology Transfer Office for he **NAVAIR ORLANDO TSD**. If either Collaborator fails to identify such counsel upon request, then such notices shall be sent to the points of contact specified above.

Article 12. SURVIVING PROVISIONS

The Articles covering Definitions, Representations and Warranties, Funding, Reports and Publications, Intellectual Property, Tangible Property, Liability, General Provisions, Modifications and Notices, and Surviving Provisions shall survive the completion, termination, or expiration of this Agreement.

Article 13. DURATION

This Agreement expires three years after its Effective Date, unless otherwise extended in writing according to the provisions of Article 11.

Article 14. SIGNATURES

For **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

I, the undersigned, am duly authorized to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002

By: M.J. Soileau

Title: Vice President, Office of Research

For the Department of the Navy:

I, the undersigned, by 15 USC 3710a and Navy regulations, am duly authorized to bind the U.S. Navy to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002.

By:

R. A. MOHLER

Title: Captain, U.S. Navy
Commanding Officer

Navy Organization: NAVAIR ORLANDO, TRAINING SYSTEMS DIVISION

This Page Intentionally Left Blank

APPENDIX A

STATEMENT OF WORK
BETWEEN
NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND
THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

NAVAIR ORLANDO TSD will:

1. Provide, for use in the UCF MURI/OPUS laboratory, source code/databases developed or enhanced by NAVAIR ORLANDO TSD as part of Navy sponsored research and development, e.g., Host Software, Image Generator Software, Weapon Tracking Software, Networking/Communications Software, Data Collection/Analysis Software, Scenario Builder Software, Rural and Urban Databases

2. Provide expertise, on a cost basis and as mutually determined to be required, on utilization of government provided software and databases for UCF MURI/OPUS Laboratory research

3. Provide, on a cost basis and as mutually determined to be required, expertise for further hardware/software/database development, testing, and training applications

4. Participate, on a cost basis, in mutually selected specific research projects conducted under the auspices of the UCF MURI/OPUS Laboratory

5. Provide access to NAVAIR ORLANDO TSD customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

3. Mutually develop processes separate from this agreement to allow exchange of funds, should that be mutually determined to be required during the period of this CRADA

NOTE 1: The parties agree that Article 9.2 Extent of Liability, will be subject to and interpreted in accordance with Florida Statutes, Section 768.28.

NOTE 2: FOREIGN ACCESS TO TECHNOLOGY

All work described in this SOW shall be subject to the following with regard to foreign access to technology.

The Collaborators agree that research findings and technology developments under this CRADA may constitute a significant enhancement to the national defense. Accordingly, access to important technology developments under this CRADA by foreign firms or institutions must be carefully controlled. These controls are in addition to, and are not intended to change or supercede, the provisions of the International Traffic in Arms Regulation (22 CFR pt. 121 et seq.) the DoD Industrial Security Regulation (DoD 5220.22-R) and the Department of Commerce Export Regulation (15 CFR pt.770 et seq.)

The non-Navy Collaborator shall provide timely notice to the government's designated Principal Investigator of any proposed transfer of research findings and technology developments under this CRADA to foreign firms or institutions. If the government determines that the transfer may have adverse consequences to the national security interests of the United States, the non-Navy Collaborator and the government shall jointly endeavor to find alternatives to the proposed transfer which obviate or mitigate potential adverse consequences of the transfer but which provide substantially equivalent benefits to the non-Navy Collaborator.

APPENDIX B

CONFIRMATORY LICENSE AGREEMENT	*1. APPLICATION FOR (Title of Invention)*
2. INVENTOR(S) AND AFFILIATION	
3. PATENT APPLICATION SERIAL NO.	*4. PATENT APPLICATION FILING DATE*
5. NAVY ACTIVITY (Name, address, point of contact)	*6. NON-NAVY ACTIVITY (Name, address, point of contact)*
7. CRADA AGREEMENT NO.	*8. DATE OF THIS AGREEMENT*

9. The Invention identified above is a "Subject Invention" under Article 7 Intellectual Property included with the CRADA identified in Box 7 between the Department of the Navy and Non-Navy Activity identified in Box 6.

This document is confirmatory of the nonexclusive, irrevocable, paid-up license to practice the identified Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the receiving party, and of all other rights acquired by the receiving party by the referenced clause.

This license is granted to

_____ *the Government*

(Select one)

_____ *Non-Navy Activity identified in Box 6*

under this CRADA in the identified Invention, Patent Application and any resulting patent.

The licensee is hereby granted an irrevocable power to inspect and make copies of the above-identified Patent Application.

_____ *ACTIVITY NAME OF LICENSOR*

_____ *SIGNATURE*

_____ *NAME (Typed or Printed)*

_____ *TITLE*

_____ *BUSINESS TELEPHONE*

University of Central Florida

Multi-Disciplinary University Research Initiative

MURI-OPUS

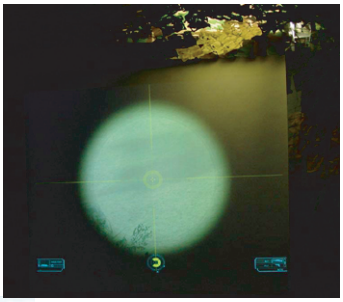
Operator Performance Under Stress

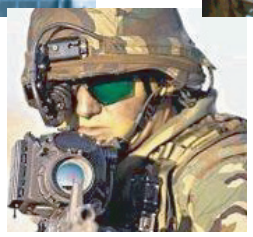
www.mit.ucf.edu

VOLUME SIX











2006 Executive Summary

Table of Contents

NCRADA

A Collaborative Partnership between:

University of Central Florida

Contact Information:
University of Central Florida • MIT² Laboratory
3100 Technology Pkwy, Suite 337 • Partnership II • Orlando, Florida 32826-0544
http://www.mit.ucf.edu

Executive Summary

Accomplishments to Date

We are approaching the end of the fifth and final year of the Multiple University Research Initiative (*MURI*) program on Operator Performance Under Stress (*OPUS*). This report represents an account of our progress to date, which contains numerous successes that have been built on previous achievements. Following the original proposal, we have elaborated upon a number of areas that remain unresolved about the influence of workload and stress on performance. Each of these issues has served as a central theme in our on-going series of 'White Papers' that have acted to guide empirical research efforts. The completed 'White Papers' are included in the present report and add to our initial model to form the basis of our theoretical foundation. Although we have detailed the progress of each discrete section of the MURI program individually, we continue to strengthen and exploit each of the theoretical, methodological, and electronic links between our respective laboratories. At the University of Central Florida, the MIT^2 research group, as will be evident, has made considerable progress. We have used our previously established research test beds to evaluate our respective theoretical propositions. From the work done so far we now know much more about variations of spatio-temporal perception under stressful conditions. We have also established that Fuzzy Signal Detection Theory (FSDT) conforms, in general, to the assumptions of traditional signal detection theory. However, preliminary evidence indicates that observers have difficulty establishing response criteria when judging fuzzy stimuli. This suggests that one problem in decision making under stress and uncertainty may be in setting a criterion for responding. Evidence for the utility of the FSDT model was established by use of more complex stimuli than in previous experiments. Specifically, ROC functions were derived using a tank discrimination task using morphed images of M1A1 and T55 tanks, and were of the same general form as observed in other experiments. Further experiments have established that forcing observers to make yes/no decisions reduced their perceptual sensitivity relative to those allowed to make tentative (i.e., rating) judgments. With regard to the latter, evidence indicates that permitting 5 categories of response yields ROC functions similar to those obtained using 100 response categories. We have also established the applicability of certain of these detection measures for technology transfer to use in the civilian realm such as medical diagnostic application to monitoring tasks. These measures promise to supplement traditional signal detection metrics in the evaluation of performance on monitoring tasks.

MURI has also examined the role of individual differences in adaptation to stress associated with perceptual and cognitive tasks. Evidence to date indicates that the personality traits of individuals influence their stress level and the coping strategies they employ to deal with the stress. One variable, attentional control, has been shown to be linked to workload and stress associated with using adaptive automation. In addition, we have found that traits such as Extraversion and Emotional Stability interact to influence how individuals react to the stress and workload of shooting.

MURI researchers have also continued their experimental investigation of the stress of sustained attention, which is relevant to current combat environments. Findings indicate that providing observers an illusion of control over the kind of vigilance task they experience improves performance, although only if the condition they receive matches their preference. Other

experimental work has established that the capacity for training with knowledge of results feedback to reduce stress symptoms depends on the dispositional pessimism of the observer.

MURI field research has been extended to examine shooting performance at a police training shooting range. This facility allows us to measure stress in realistic environments similar to those of modern military soldiers in broken urban environments. An initial study confirmed the workload-performance association/dissociation pattern predicted by the Hancock and Warm model in the context of a real-world shooting task with professional police officers. Of particular relevance is that these data were inconsistent with the traditional 'inverted U' explanation of performance, further indicating that this older representation of stress and performance does not reflect real-world data. The results of this study also confirmed the importance of considering the subscales of perceived workload measures, and underscore the multidimensionality of the construct itself. Through a Cooperative Research and Development Agreement (*CRADA*) with the Naval Air Warfare Center, Training systems Division, Orlando, the MIT^2 laboratory has created, refined, and now exploited the Small Arms Simulation Test Bed (SAST) facility. We have conducted experimental investigations of cognitive changes that occur in a stressful target acquisition task. Findings indicated performance on an auditory secondary task, in which the individual must process information regarding the identification and location of other friendly forces, is disrupted by concurrent target acquisition demands. Although shooting accuracy was not significantly impacted by the presence of a secondary task, response time was slower at the highest demand level, indicating a cost for maintaining accuracy under high workload conditions. Further, the performance decrement was observed only on trials in which a target was present.

The MURI group has expanded its capability for integrating physiological measurement into experimental investigations. These include heart rate variability and eye tracking, and we anticipate expanding our systems to include other forms of physiological response. Physiological data for the SAST experiments described above are currently in the analysis stage.

The *VIKING* experimental environment, originally developed at the University of Minnesota, has been used to assess performance on tasks that impose information processing demands equivalent to those faced by soldiers in combat environments. Currently the MIT^2 Laboratory is using the recently developed ARES battery to assess the changes in cognitive performance of soldiers in the field during simulated combat, finding that many soldiers tolerate such field conditions with minimal impairment of cognitive performance if they know what is expected but that high demand events are still associated with cognitive performance decrement. In addition, it appears that different forms of cognitive task are not affected equally in the context of combat. Cases were observed in that study that clearly showed a precipitous decline in adaptive function at the extremes of stress, as predicted by the Hancock and Warm (1989) model. The decrement in cognitive function was replicated in a sample of participants in a Survival Evasion, Resistance, and Escape (SERE) exercise, with the more complex cognitive tasks (e.g., logical reasoning) showing the greatest susceptibility to the effects of sustained stress. In addition, performance immediately after the exercise showed a return to normal (pre-exercise) levels, but only for a short period of time. The field work using the ARES has established the utility of the battery of tasks for measuring the effects of stress in operational settings (e.g., field training exercises) on perceptual and cognitive processes.

As well as these on-going empirical efforts, researchers in the MIT^2 laboratory have completed meta-analytic reviews of the effects of temperature, vibration, and noise on human performance. A fourth meta-analytic study of the effect of time pressure on performance is nearing completion. These efforts are geared toward enhancing IMPRINT capacity for prediction of performance under stress. Each of these stressors impair performance, but in the case of thermal stress, the impact depends on the range of temperature, as well as the nature of the task and the duration of exposure. It is likely that intensity, duration, and task type also moderate the relation of vibration to performance, but there were insufficient studies in the published literature to evaluate these hypotheses. More fine grained analysis of vibration and performance will require programmatic experimental work by future researchers.

The Team Performance Laboratory (*TPL*) at the University of Central Florida has investigated adaptability of teams under time pressure and high workload, and is currently using a military-based task (*Flashpoint*) to further investigate the impact of stress on team processes. In particular, derived evidence indicates that the quality of team communication and performance varies as a function of team type (e.g., command and control versus a reconnaissance team). TPL researchers have also found that time pressure impairs communication among team members (breaking the closed-loop communication pattern). Recent evidence indicates that teamwork behaviors (i.e., communication) and team performance under stress are influenced by task interdependency. These results indicate that teams should not be treated as equivalent and that team training should incorporate differences in team type and in the degree of task interdependency. In addition to these efforts, TPL researchers have made advances on the development of metrics to assess team performance under conditions of high stress and workload. It is anticipated that these can be transitioned to further research to implement the empirical and theoretical advances of this MURI to developing training regimens for teams, as well as transitions to civilian domains. Toward that end, TPL researchers have developed 'best practices' guidelines for team training.

MURI researchers at Linköping University (LiU), (formerly at Kansas State University) have investigated the effects of trust on team decision-making, particularly, the identification and mitigation of cognitive decrements due to remote command and control under realistic simulations of battlefield conditions. These efforts have resulted in the completion of a viable research paradigm for the investigation of trust and team problem-solving and represents one of the myriad examples of cross Laboratory technology transfer and protocol integration. The LiU group has conducted several simulated combat environment experiments on the effects of social interaction on performance. These experiments document a cognitive decrement directly attributable to remote command and control. Military simulations using paintball revealed that when leaders engage in remote command and control there is a decrement in soldiers' trust for the leader. Associated with this loss of trust is compromised performance. Further experimentation has indicated that use of vibrotactile stimulation to issue commands remotely can compensate for hesitation to commands from remote commanders. They have also established that the use of information and communication technology for emergency management can lead to performance decrements if there is a mismatch in its use between collocated training and distributed emergency management operations. They have also extended their decision making research to the investigation of the effects of cultural differences on decision making efficiency in multicultural command-and-control centers. Their research indicates that culturally-based differences in planning and leadership style can impair decision making efficiency in these contexts.

MURI researchers in the Cognitive Science Laboratory at the George Mason University (formerly at Catholic University of America) have completed a sequence of studies on the impact of automation on performance in decision-making tasks of the kind typically faced by military personnel in command and control settings. They found that as a result of complacency, automation with high (but not perfect) reliability does not always best support performance. They also found that automation unreliability has a greater performance cost for decision automation than for information automation. In addition, they continue to explore the effectiveness of adaptive automation using real-time measurement of workload under conditions of sustained workload and perceived stress. They are currently working on examining blood flow as an index of cognitive workload using the Transcranial Doppler Sonography System (TCDS). System sensitivity to blood flow changes in the left and right, middle cerebral arteries is being examined in a working memory task involving storage and recall of digits and in the tracking sub-task of the Multi-Attribute Task Battery. Researchers at the University of Cincinnati have conducted experiments demonstrating that the decline in sustained attention performance over time is accompanied by a parallel decline in cerebral bloodflow in the right hemisphere. In addition, they have initial evidence that transcranial oximetry (TCCO) measures of blood flow and oxygenation levels in the brain may also provide a useful physiological measure of sustained attention. In a recent experiment investigating the effect of type of automation support on decision making it was found that performance (both speed and accuracy) in automated tasks improved when changes to information automation were adaptive, and that changing automation type in the middle of a task reduced the cost associated with automated decision aid. Thus, response time under manual control was better after information automation. These results indicated that adaptive changes in automation type during a task can reduce the costs associated with unreliable automation. In particular, changes from recommendation to listing automation formats improves both accuracy and response time, while change from decision recommendation to manual improved response time. Thus, providing some information automation during a session before switching to decision automation can facilitate performance. GMU researchers are currently working to complete experimental investigation of the potential for adaptive warning systems to mitigate the performance decrements associated with task-induced fatigue

These researchers have also developed methodology for enabling flexible supervisory control that can be delegated to automation, and they have demonstrated how these methods can be applied to UAV control.

In respect of all these achievements, MURI has supported over thirty researchers at multiple universities, including five post-doctoral researchers and over twenty graduate student researchers across all participating members. The participating universities are University of Central Florida, University of Minnesota, Minnesota State University at Mankato, Kansas State University, Linköping University, University of Cincinnati, Catholic University of America, and George Mason University.

Contributions to Scientific Understanding

We have made several fundamental steps in scientific understanding concerning stress effects. We have begun to elucidate the effects of diverse sources of real-world stress on spatio-temporal perception. Using stress sources that will impact further development of the IMPRINT model, we have seen confirmation of our hypotheses on the 'narrowing' of the perceptual world of the stressed operator. These efforts have burgeoned into display design guidelines for such operations

and in modifications to the seminal multiple resource model. We have also made progress in development of stress and workload theory by refining the fundamental questions to be resolved regarding the cognitive resources construct that underlies current stress theory. These efforts, which resulted in two book chapters currently in press, will help guide future research on stress and workload associated with human performance. We have made substantive theoretical progress on the question of adaptive interfaces for stressful operations. In particular, we have presented a number of critical insights about the efficacy of design and the possibility of optimal design. These developments have helped characterize and strengthen currently tested empirical propositions that promise to save significant dollars against a vision of endless, iterative search for design improvement. Our progress has been facilitated by naturalistic field observations and evaluations. The opportunity to conduct real-world test and evaluation on stressed soldiers and police officers accords with our theoretical proposition that a fully articulated theory can only emerge from an understanding of actual operations. Our experiments on combined physical and cognitive effort have resulted in a surprising and important discovery. We call this the 'table-top' tolerance limit phenomenon. This phenomenon indicates that even when operational stress levels are extremely high, the performance remains relatively constant to a point of sudden incipient failure. This finding has significant implications for military operations. Identifying symptomatic characteristics at this "edge of failure" (e.g., progressive increase in performance variability) can provide a predictive marker that enables the identification of impaired soldiers. This may well prove to be a crucial theoretical and practical discovery. Our empirical investigations of workload and shooting performance have enhanced the understanding of the workload-performance relations in real-world tasks, and how personality differences affect these relations. A manuscript based on this work has been submitted for publication and is contained in this report. These observations have crucial importance for both military personnel selection and military personnel training. These results support our contention that consideration of individual differences is an important component of research on stress and performance. We are also now beginning to distill ways in which team performance under stress is not simply the product of individual contributions. In particular, we have shown how workload sharing and task division and reallocation by team units acts to diffuse unacceptably high levels of demand, and the importance of task interdependence on team communication and performance. We are using this insight as a basis for designs in adaptive interfaces for stress in which demand may be shared between exposed soldiers and support personnel in remote locations. Initial developments in our collective work demonstrate how direct monitoring of operator state may be mapped into task resolution, well beyond the traditional conception of 'adaptive' interfaces. These new, neuro-ergonomic insights present the next generation level of conceptions for human-machine interaction. We have also found that stress influences mutual trust levels in combined decision-making conditions, and that tactile stimulation can mitigate this problem. Finally, empirical research on Fuzzy Signal Detection theory has supported the contention that this model can be employed to better understand and quantify detection performance when the stimulus dimensions inherently contain high levels of uncertainty. However, constraining operator response to yes/no decisions can reduce sensitivity when the stimuli to be inspected are a fuzzy set. These represent only a selection of insights derived from the contained work, others of which are articulated in the specific papers.

Research for the Remainder of the Project

The meta-analytic work on the performance effects of time pressure will be completed, as will extension of noise effects to subjective response. This latter effort will permit meta-analytic comparison of associations and dissociations between performance and subjective state. In regard

to the relation between the spatial and temporal task dimensions that form the base axes of the Hancock and Warm (1989) model, further research will explore the distortion of spatial and temporal perception under stress, and in particular the relation between these fundamental dimensions. We also are in the process of preparing a text to be published by Ashgate that will describe the research programs funded under this MURI. This text will serve as both summary of research and contribution to improvement in stress theory, and as a summary of tools and methodologies that can be used to assess soldier performance under stress from the team level down to the physiological response of the individual soldier.

Table of Contents

MURI Divisions

University of Central Florida (Peter Hancock)

Supported Work

Incorporating Individual Differences into Adaptive Automation *Presentation* 113th Annual Convention of the American Psychological Association, Washington, DC, August, 2005 *Oron-Gilad, T., Thropp, J.E., Szalma, J.L., and Hancock, P.A. (2005)*

The Driving Simulator as a Workload Research Tool *Presentation* Driving Simulator Conference North America, December 2005 *Morgan, J.F., Oron-Gilad, T., & Hancock, P.A. (2005).*

The Driving Simulator as a Workload Research Tool *Proceedings* Driving Simulator Conference North America, December 2005 *Morgan, J.F., Oron-Gilad, T., & Hancock, P.A. (2005).*

Application of Fuzzy Signal Detection Theory to the Discrimination of Morphed Tank Images *Presentation* Human Factors and Ergonomics Society 50th Meeting *Szalma, J.L., Oron-Gilad, T., Saxton, B.M., & Hancock, P.A. (2006).* In press.

Performance, Workload, and Stress in Vigilance: The Power of Choice *Presentation* Human Factors and Ergonomics Society 50th Meeting *Szalma, J.L. & Hancock, P.A. (2006).* In Press.

Low and High Levels of Attentional Control in Adaptive Automation *Poster* Human Factors and Ergonomics Society 50th Meeting *Thropp, J., Oron-Gilad, T., Szalma, J.L., & Hancock, P.A. (2006).* In press.

Development of an Operational Setting Specific Field Cognitive Assessment Procedure *Presentation* Human Factors and Ergonomics Society 50th Meeting *Harris, W.C., Reeves, D.L., Elsmore, T.F., and Hancock, P.A. (2006).* In Press.

Application of Fuzzy Signal Detection Theory to Threat Detection in Extreme Environments *Poster* Human Performance in Extreme Environments 4th Annual Meeting *Hancock, P.A., Szalma, J.L., Koltko-Rivera, M.E., & Vincenzi, D.A. (2006)* In Press.

Road Environment and Driver Fatigue Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design *Oron-Gilad, T., & Hancock, P.A. 2005* 318-324.

Police Firearms Training: Individual Differences in Performance, Workload, and Stress *Refereed Publication* Szalma, J.L., Oron-Gilad, T., Stafford, S.C., and Hancock, P.A. 2006 Submitted.

Training for Vigilance: Using Predictive Power to Evaluate Feedback Effectiveness *Refereed Publication* Szalma. J.L., Hancock, P.A., Warm, J.S., Dember, W.N., & Parsons, K.S. (2006) In press..

Training for Vigilance: The Effect of Knowledge of Result Format and Dispositional Optimism and Pessimism on Performance and Stress *Referred Publication* Szalma, J.L., Hancock, P.A., Dember, W.N., & Warm, J.S. (2006), *The British Psychological Society, 97*, 115-135.

Cognitive Changes in the Infantry Field Environment *Presentation* ANAM Sports Medicine Battery Working Group meeting, Center for Cognitive Neuroscience, National Rehabilitation Hospital, Washington, DC *Harris, W.C., & Hancock, P.A. (2006).*

Vibrotacile Guidance Cues for Target Acquisition *Referred Publication* Oron-Gilad, T., Downs, J.L., Gilson, R.D., & Hancock, P.A. *IEE SMC Part C (2006)* In press.

An examination of teamwork communication across teams *Referred Publicatio*n Priest, H. A., Burke, C. S., Guthrie, J., Salas, E. & Bowers, C. A. (2006). Are all teams created equal?. *Human Factors.* under revision

Work teams in organizations: A historical reflection and lessons learned *Book Chapter* Salas, E., Priest, H. A., Stagl, K. C., Sims, D. E., & Burke, C. S. (2006). In L. L. Koppes (Ed.), *Historical Perspectives in Industrial and Organizational Psychology*. Mahwah, NJ: LEA.

Assessment of Team Learning *Book Chapter* Stagl, K. C., Salas, E., & Day, D. V. (in press). In V. Sessa & M. London (Eds.), *Group learning* (pp.). Mahwah, NJ: Lawrence Erlbaum and Associates.

Best Practices in Cross Training Teams *Book Chapter* Stagl, K. C., Salas, E., & Fiore, S. M. (in press). In D. A. Nembhard (Ed.), *Workforce cross training handbook* (pp.). Boca Raton, FL: CRC Press.

Understanding Performance, Workload, and Stress: An Individual Differences Perspective *Abstract* Human Factors Issues in Complex System Performance *Szalma, J.L. & Hancock, P.A. (2006) Accepted.*

Fuzzy Signal Detection Theory and Human Performance: A Review of Empirical Evidence for Model Validity *Abstract* Human Factors Issues in Complex System Performance *Szalma, J.L. & Hancock, P.A. (2006) Accepted.*

Facilitated Work

A Meta-Analysis of Periodic Noise Stress on Human Performance *Poster* Fourth Annual Meeting of the Society for Human Performance in Extreme Environments. *Ross, J.M., Conway, G.E., Szalma, J.L., Saxton, B., Braczyk, A. & Hancock, P.A (2006).* In Press.

A Meta-Analysis of Performance Response Under Thermal Stress *Refereed Publication* Hancock, P.A., Ross, J.M., & J.L. Szalma (2006) *Human Factors*. In press.

A Meta-Analysis of Performance Response Under Thermal Stress *Poster* Human Factors and Ergonomics Society 50th Meeting *Ross, J.M., Szalma, J.L., & Hancock, P.A. (2006)* In press.

A Quantitative meta-analytic examination of whole-body vibration effects on human performance. *Referred Publication* Conway. G.E., Szalma, J.L., & Hancock, P.A. *Ergonomics* (2006) In Press.

The effects of whole-body vibration effects on human performance: A meta-analytic examination *Poster* Human Factors and Ergonomics Society 50th Meeting Conway, G., Szalma, J., Saxton, B., Ross, J., and Hancock, P. (2006) In Press.

Stress and neuroergonomics. *Referred Publicatio*n Hancock, P.A., & Szalma, J.L. (2006). In: R. Parasuraman and M. Rizzo (Eds.), *Neuroergonomics: The brain at work*. Oxford: Oxford University Press. (pp 195-206)

A Meta-Analysis Of Aperiodic Noise Stress On Human Performance. *Abstract* Army Sciences Conference *Saxton, B., Ross, J.M., Braczyk, A., Conway, G., Szalma, S.L., & Hancock, P.A.* November 2006 Orlando, FL.

Communicating Through The Use Of Vibrotactile Displays For Dismounted And Mounted Soldiers. *Abstract* Army Sciences Conference *Merlo, J.L., Terrence, P.I., Stafford, S., Gilson, R., Hancock, P.A., Redden, E.S., Krausman, A., Pettitt, R., and White, T.L.* November 2006 Orlando, Florida.

Linköping Institute of Technology (Kip Smith)

Supported Work

A case study of information and communication technology in emergency management training. *Refereed Publication* Woltjer, R., Lindgren, I., & Smith, K. (2006) *International Journal of Emergency Management* In Press.

A brain imaging study of procedural choice. *Refereed Publication* Rustichini, A., Dickhaut, J., Ghirardato, P., Smith, K., & Pardo, J. V. (2005). *Games and Economic Behavior, 52 (2)*, 257 - 282.

Remote command and control compromises soldiers' trust in their leaders. *Proceedings* Human Factors and Ergonomics Society 50th Meeting *Liberg, M., & Smith, K. (2006).* In press.

Using microworlds to understand cultural influences on distributed collaborative decision making in C2 settings. *Proceedings* 11th Annual International Command and Control Research and Technology Symposium. Cambridge UK. *Lindgren, I., & Smith, K.* (2006). In Press.

Facilitated Work

Economics and emotion: Institutions matter. *Refereed Publication* Smith., K., & Dickhaut, J. (2005) *Games and Economic Behavior, 52 (2),* 316 – 335.

Progress Report – 09/2005

Progress Report – MURI Final 07/2006

George Mason University (Raja Parasuraman)

Supported Work

Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Referred Publication* Miller, C., & Parasuraman, R. (2006). *Human Factors, 48.*

Effects of imperfect automation on decision-making in a simulated command and control task. *Referred Publication* Rovira, E., McGarry, K., & Parasuraman, R. (2006). *Human Factors, 48.*

Introduction to neuroergonomics. *Book Chapter* Parasuraman, R., & Rizzo, M. (2006). In R. Parasuraman, & M. Rizzo (Eds.) *Neuroergonomics: The Brain at Work.* (pp. 1-12). New York: Oxford University Press.

Future prospects for neuroergonomics. *Book Chapter* Rizzo, M., & Parasuraman, R. (2006). In R. Parasuraman, & M. Rizzo (Eds.) *Neuroergonomics: The Brain at Work.* New York: Oxford University Press.

Cerebral hemodynamics and vigilance. *Book Chapter* Warm. J. S., & Parasuraman, R. (2006). In R. Parasuraman & M. Rizzo (Eds.) *Neuroergonomics: The Brain at Work.* (pp. 146-158). New York: Oxford University Press.

Prevention of rear-end crashes in drivers with task-induced fatigue through the use of auditory collision avoidance warnings. Proceedings May, J., Baldwin, C. L., & Parasuraman, R. (2006, October). *In Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society,* San Francisco, CA.

Progress Report – 07/2005

Progress Report – 10/2005

Progress Report – 09/2006

Incorporating Individual Differences into Adaptive Automation

Tal Oron-Gilad, Ph.D
Jennifer E. Thropp
James L. Szalma, Ph.D.
Peter A. Hancock, Ph.D.



MIT² Laboratory

Institute for Simulation & Training
and Department of Psychology
University of Central Florida, Orlando FL

APA Convention 2005
Washington, DC, Aug 18-21

Focus and Process

Fabricators of Technology

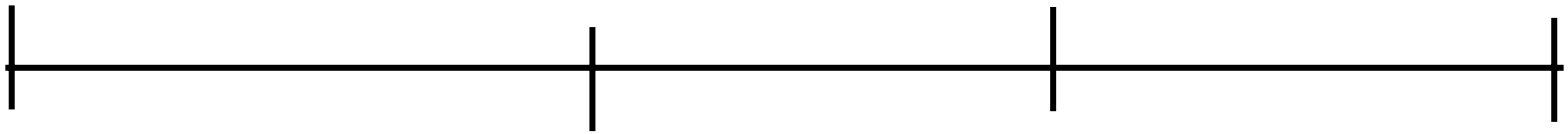
Self-Idiographic

Collective Nomothetic

Self-Nomothetic

Collective-Idiographic

Self (Non-Optimal) | Collective | Collective | *Individual*



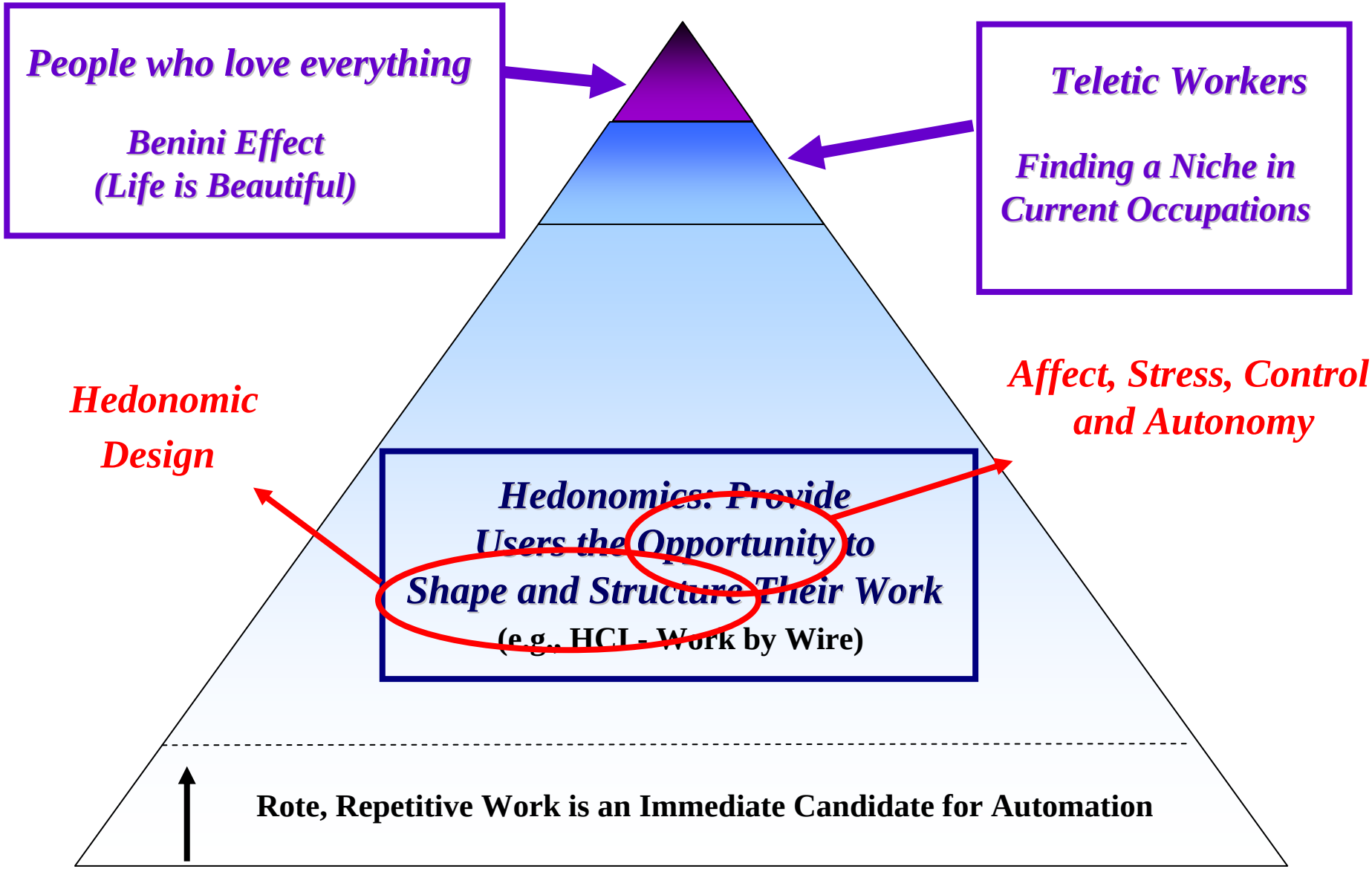
Customization | *One size fits all* | *Adaptation* | *Individuation*

Cottage Industry | Mass Production | Usability | Hedonomics

Users of Technology



Time



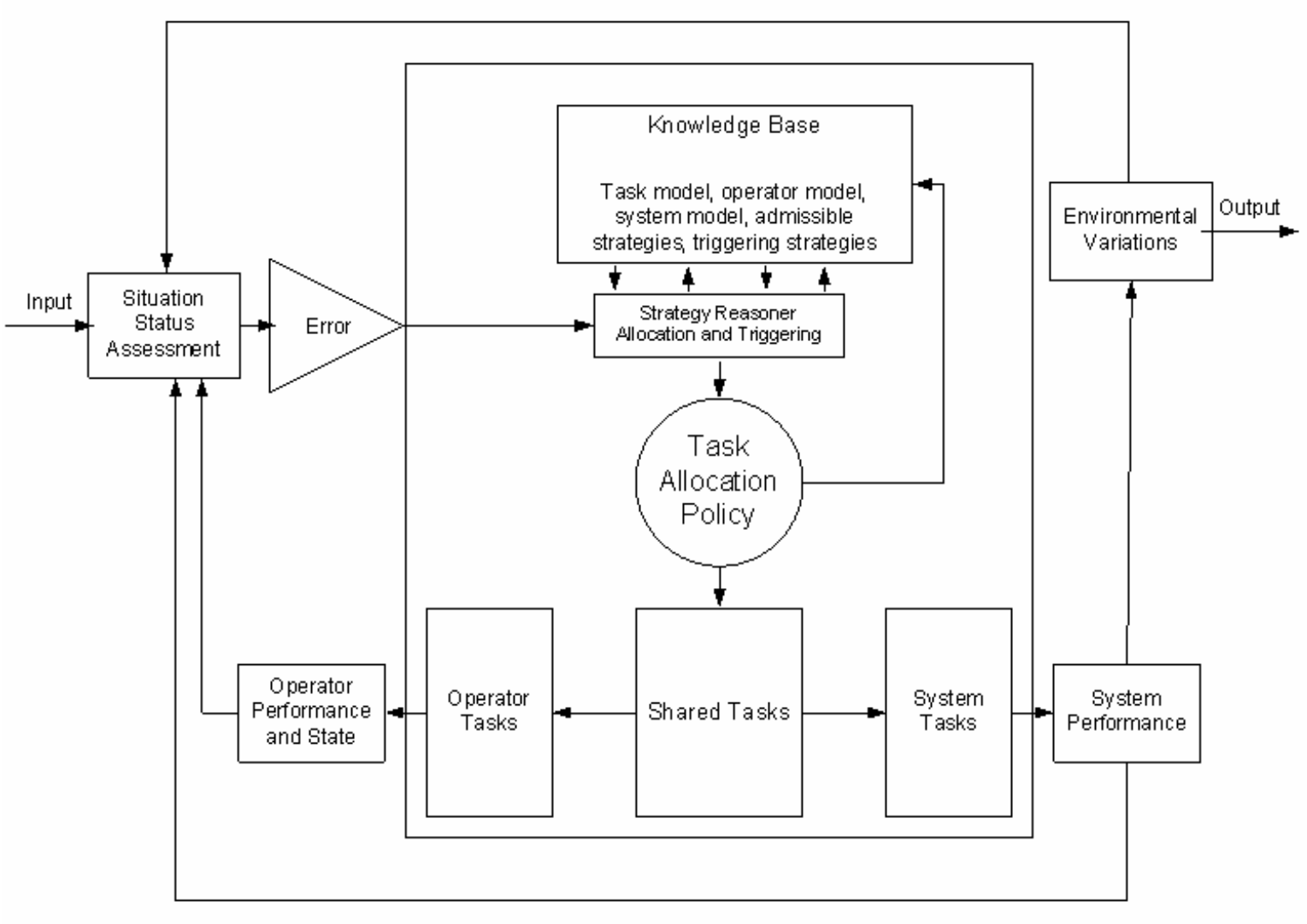
Seligman, M.E.P. (2002). Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment. NY: Simon & Schuster.

Cskikszentmihalyi M. (1990). Flow: The psychology of optimal experience. Harper & Row Publishers

Why is it the right time

- Feasible, technology wise
- ID account for more variability than design (Egan, 1988). SA accounts for 88% of human error in aviation (Ensley, 1995)
- Less selection
- Reduce training costs

What is Adaptive Automation?



Oron-Gilad, Thropp, Szalma & Hancock. Incorporating ID into adaptive automation.

Levels of Automation (after Sheridan)

LEVELS OF AUTOMATION OF DECISION AND ACTION SELECTION

HIGH	10. The computer decides everything, acts autonomously, ignoring the human.
	9. informs the human only if it, the computer, decides to
	8. informs the human only if asked, or
	7. executes automatically, then necessarily informs the human, and
	6. allows the human a restricted time to veto before automatic execution, or
	5. executes that suggestion if the human approves, or
	4. suggests one alternative
	3. narrows the selection down to a few, or
	2. The computer offers a complete set of decision/action alternatives, or
LOW	1. The computer offers no assistance: human must take all decisions and actions.

Oron-Gilad, Thropp, Szalma & Hancock. Incorporating ID into adaptive automation.

Management by Consent MBC, Management by Exception MBE

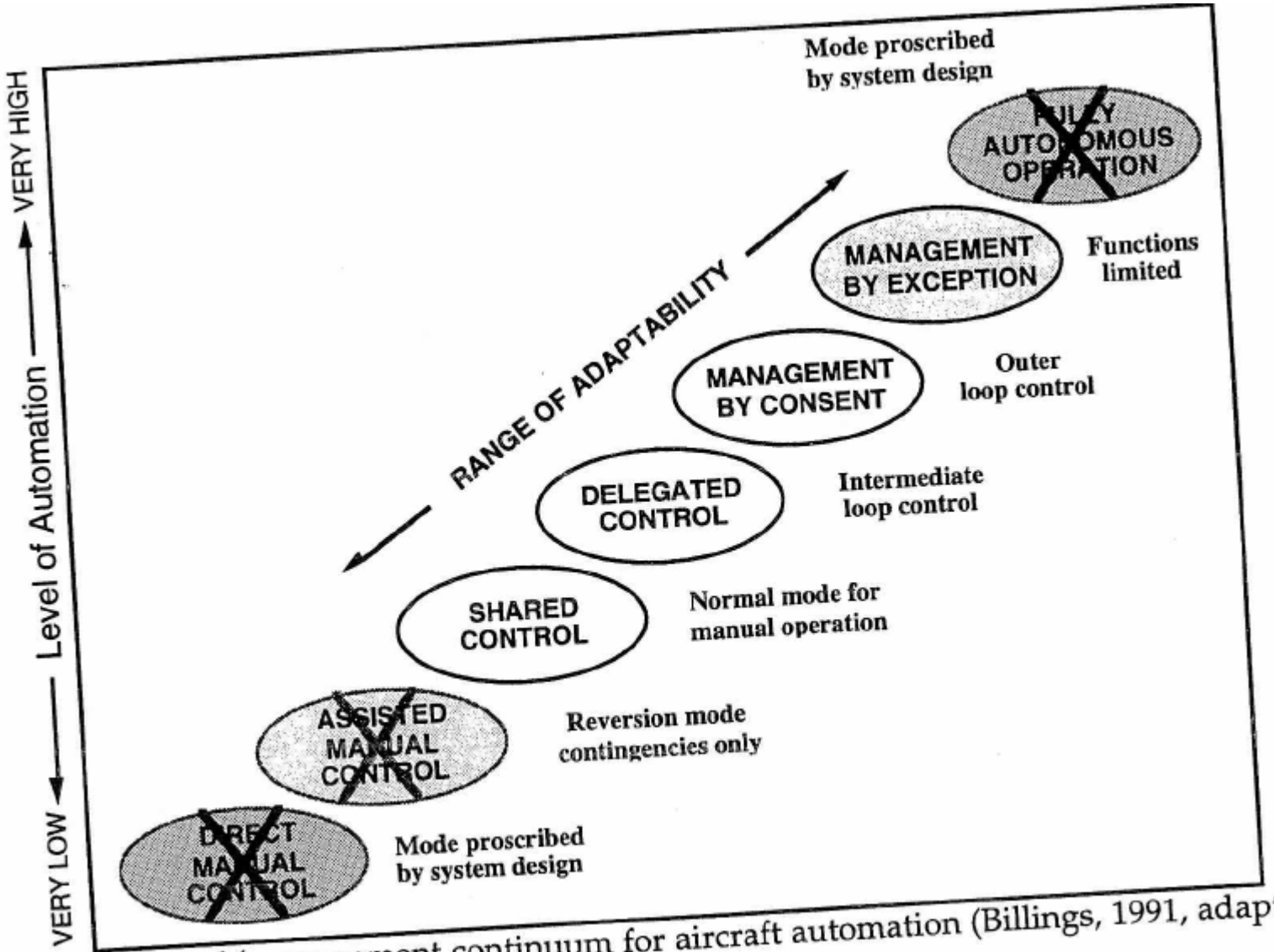
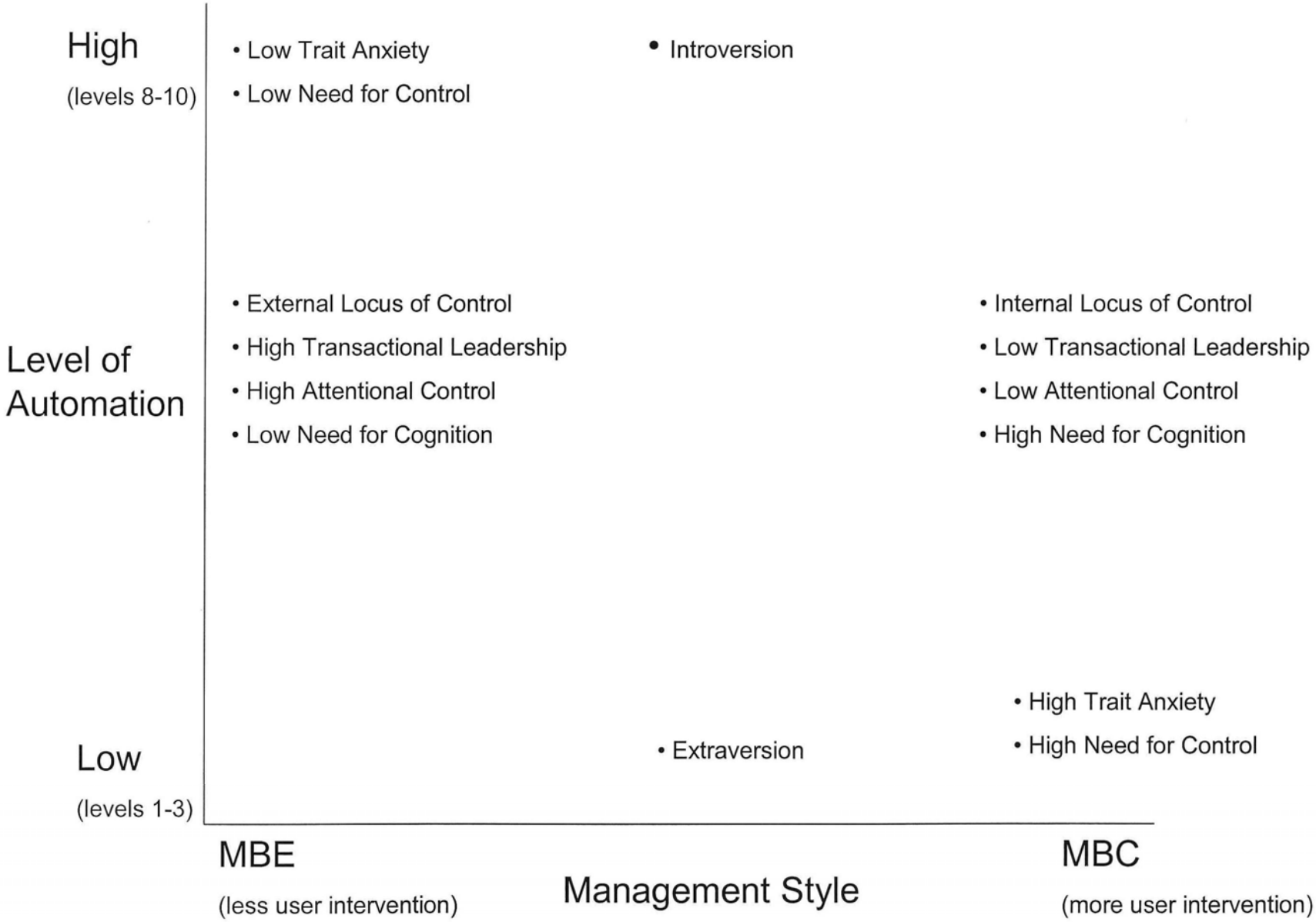


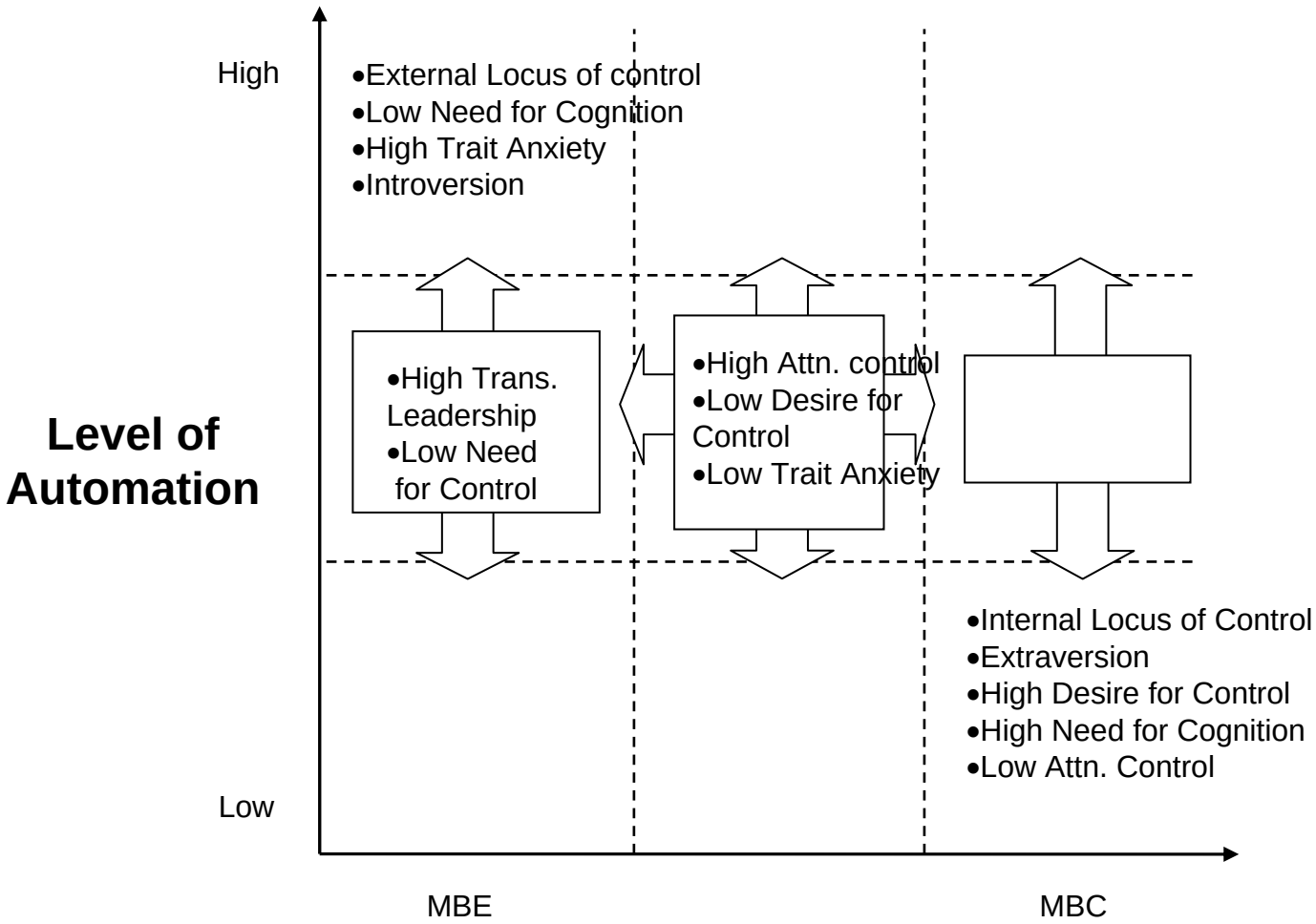
Figure. 6: Control/management continuum for aircraft automation (Billings, 1991, adapted).

Billings & Woods, 1996



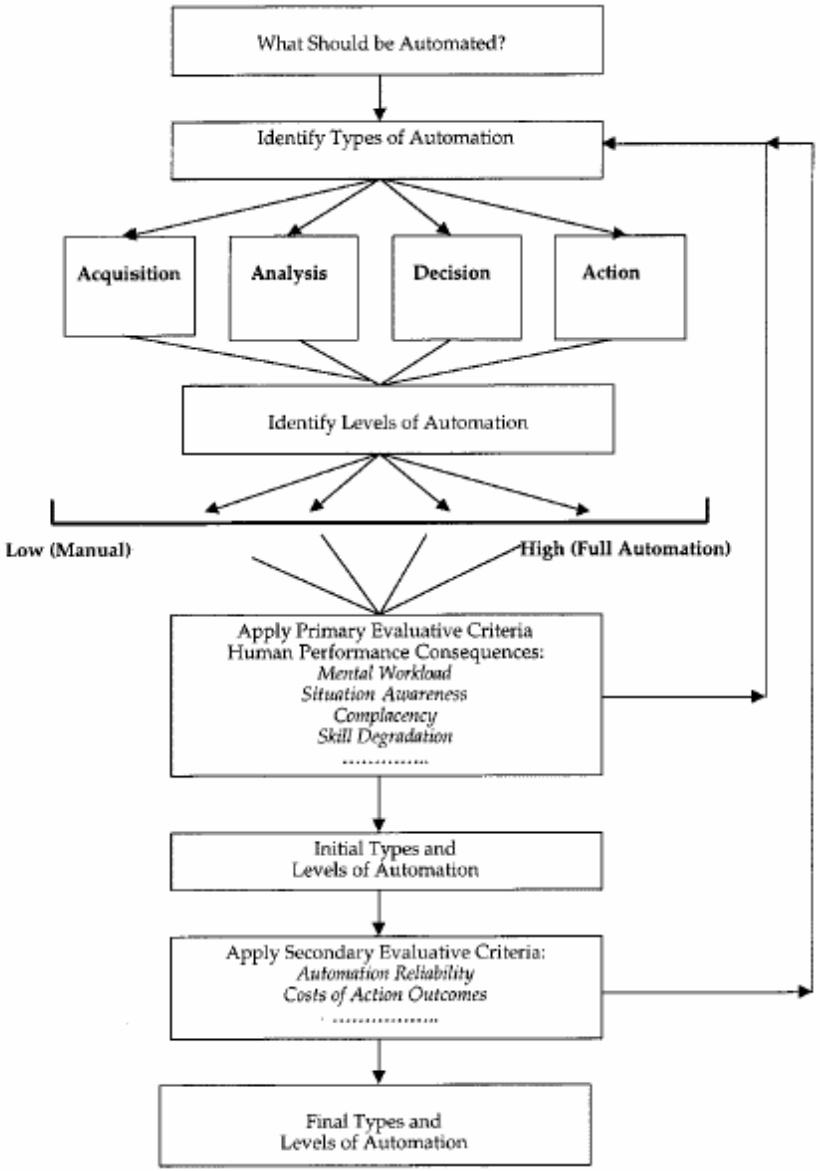
Oron-Gilad, Thropp, Szalma & Hancock. Incorporating ID into adaptive automation.

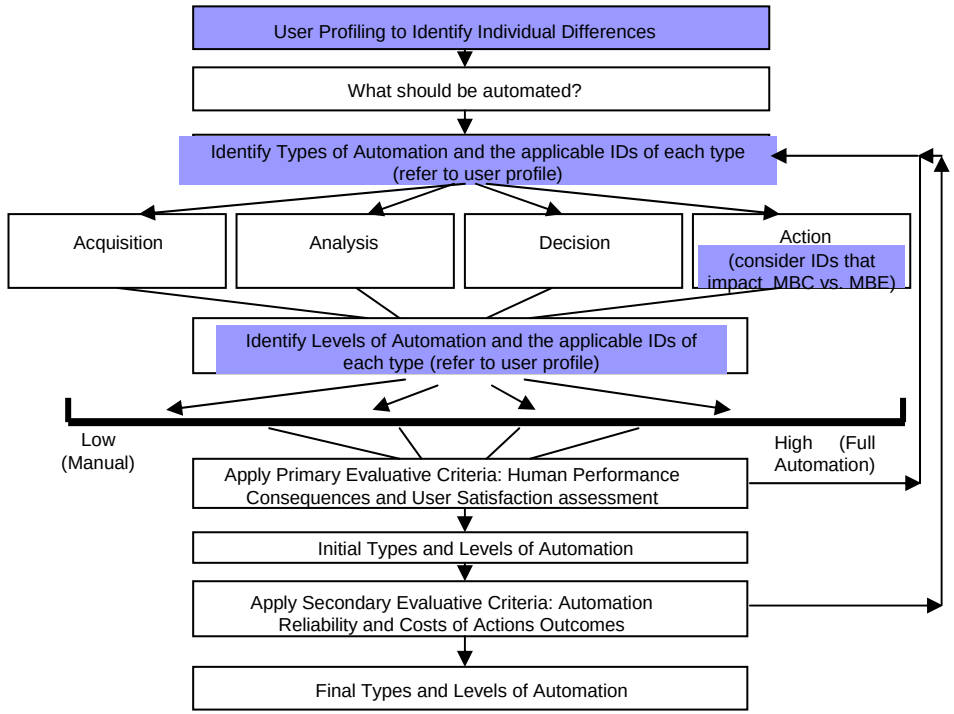
ID in the Matrix



What should be automated

(Parasuraman,Sheridan, Wickens, 2000)





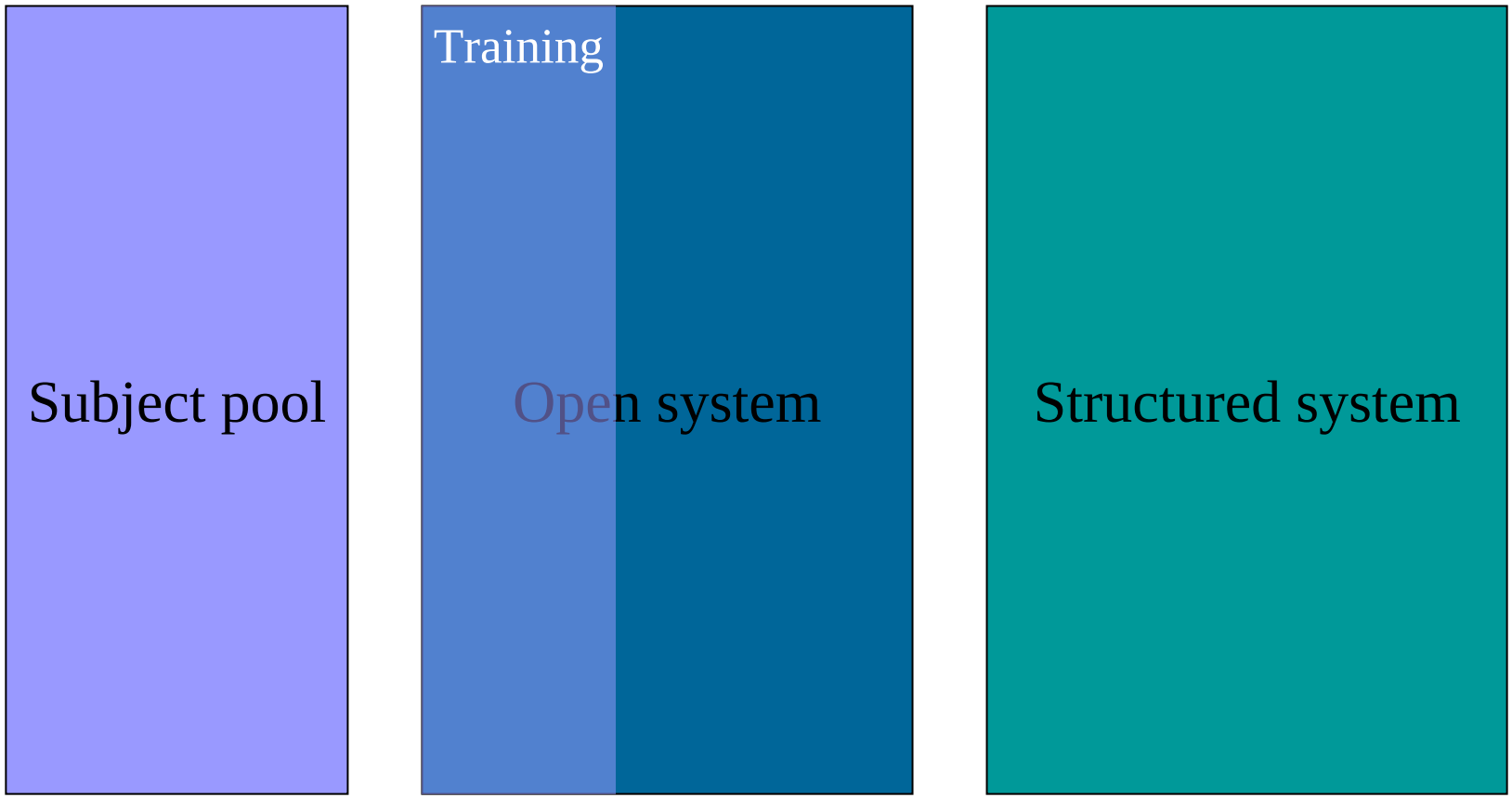
Its not just the big five

Traits	Relevance	Robust Measure	Empirical Support		Relevancy	
			General	Automation	Mgmt Style	LOAs
Attentional Control	How well can MBC be implemented to prompt users to allocate their attention to important aspects of the task/display	Low	Med	Low	✓	✓
Trait Anxiety	task-specific performance effects, feedback Implications, may provide significant predictive ability.	High	Med	Low	✓	✓
Desire for Control	Direct relationship with control of the automated system, in terms of both performance and user preference. Should be examined in regards to adaptive systems. Important to assess its performance implications.	Med	Med	Med	✓	✓

Its not just the big five (Cntd.)

Traits	Relevance	Robust Measure	Empirical Support		Relevancy	
			General	Automation	Mgmt Style	LOAs
Need for Cognition	Implications for information presentation. Level of Interactivity and SA relationship.	Med	Low	Low	✓	✓
Extraversion	Task-specific performance effects, feedback implications. Important to examine its effects on response behavior.	High	High	Low	✓	✓
Locus of Control	important to determine the degree to which external locus of control is resistant to external control exerted by the automation.	High	Med	Med	✓	✓
Leadership Style	Has only been discussed in terms of human-human interaction. Rooted in the analogy between the 'human-human'& 'human-machine' interaction styles.	Low	Med	Low	✓	

Experimental Approach



Acknowledgments

This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621, Dr. Elmar Schmeisser, *Technical Monitor*, P.A. Hancock, *Principal Investigator* .

This work was supported in part by the Army Research Laboratory through the Micro-Analysis and Design CTA Grant DAAD19-01C0065, Michael Barnes, *Technical Monitor*, P.A. Hancock, *Principal Investigator*.

The views expressed in this work are those of the authors and do not necessarily reflect official Army or Department of Defense policy.

The Driving Simulator as a Workload Research Tool

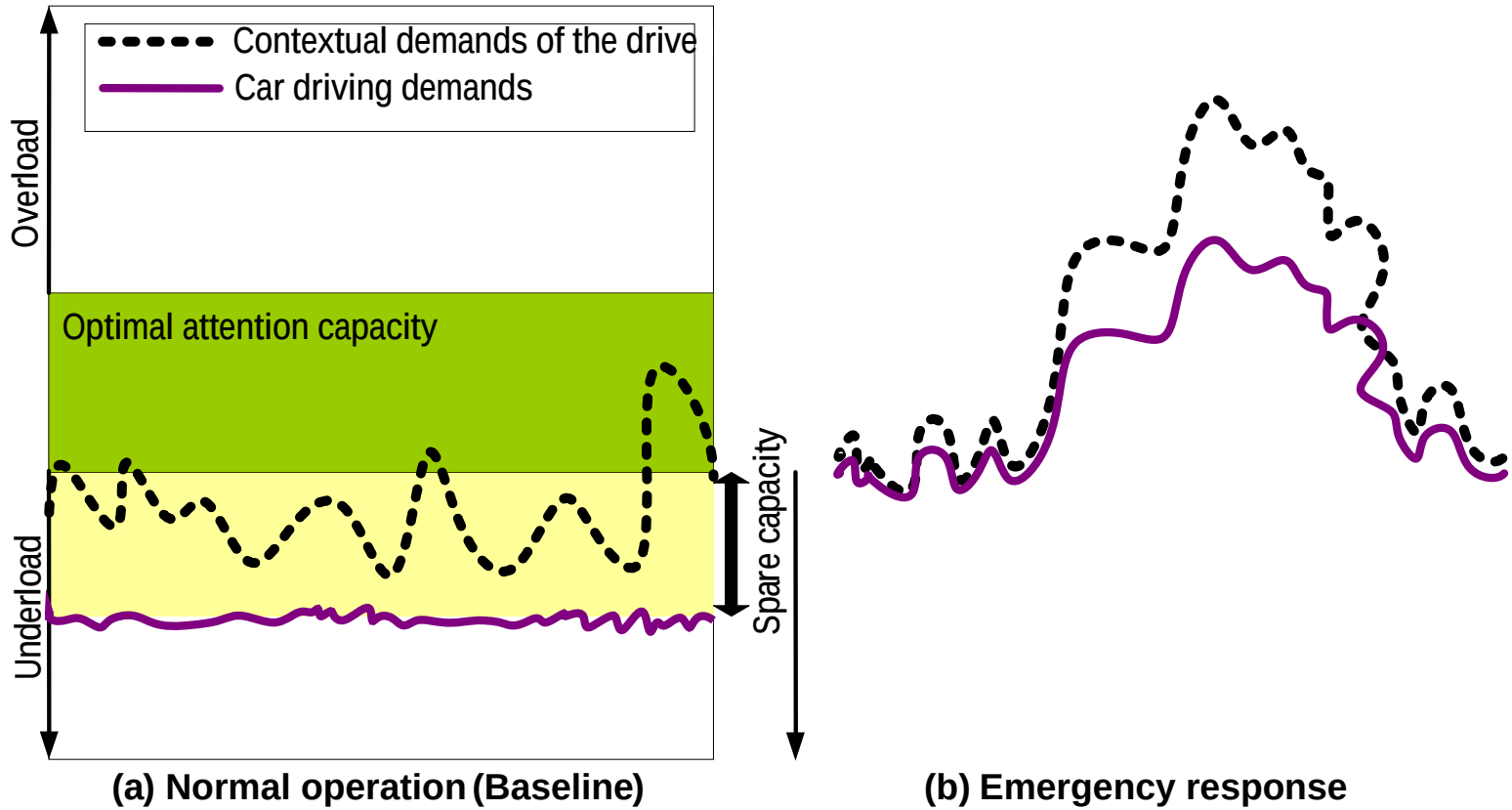
Minds in Technology-Machines in Thought
MIT² Laboratory
University of Central Florida

J.F. Morgan,
T. Oron-Gilad,
& P.A. Hancock

Introduction

- Driving is a variable workload task
  - Driving affects workload
  - Simulators affect workload
  - Workload is a difficult construct to measure
- How to separate those becomes the issue

Driving as a Variable Workload Task

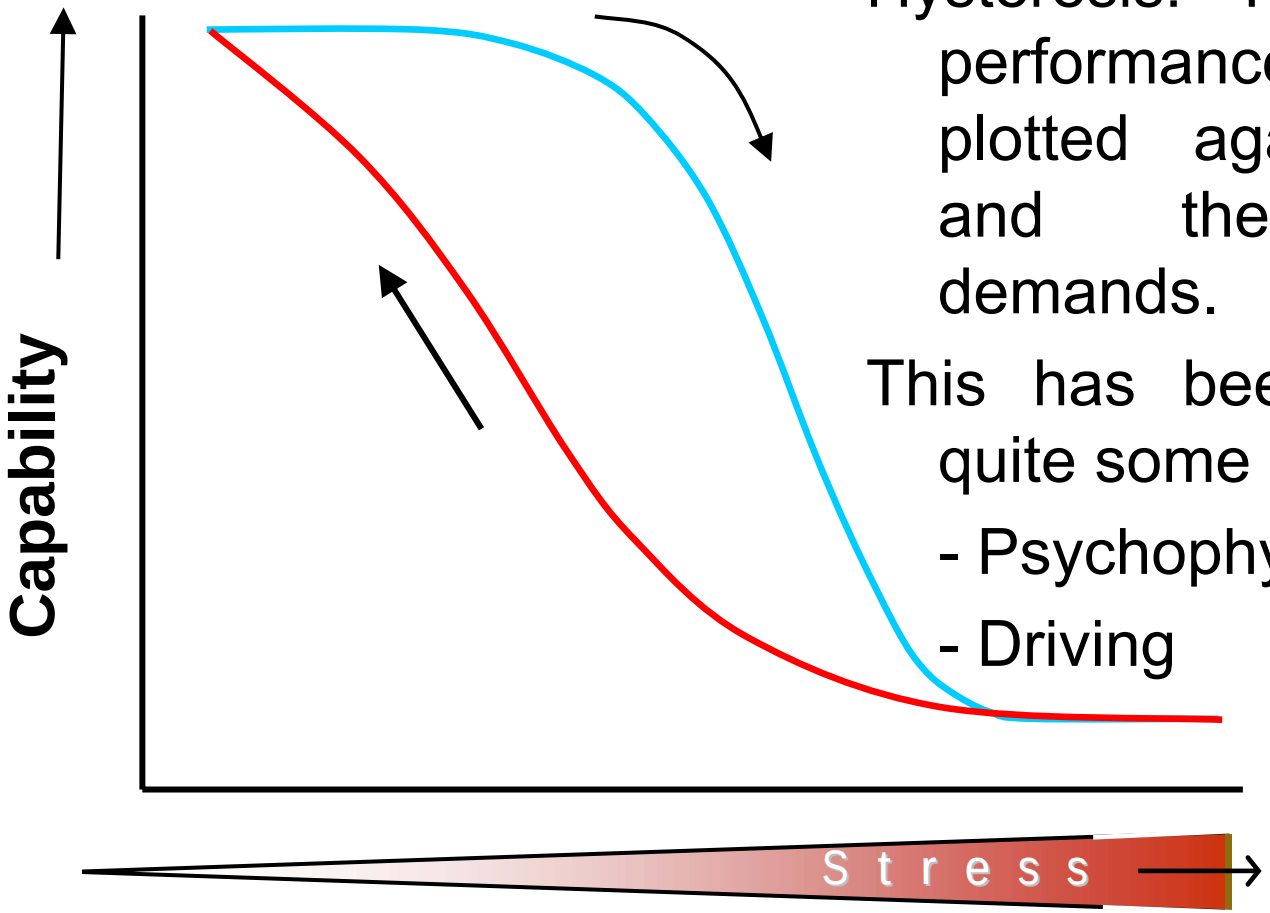


(a) Normal operation (Baseline)

(b) Emergency response

Driving is dynamic. Contextual demands of a drive may vary rapidly.

Hysteresis



Hysteresis: The shape of a performance curve when plotted against increasing and then decreasing demands.

This has been observed for quite some time in humans.

- Psychophysical research
- Driving

Research Objective

Task where stress/workload of the task may be separated from stress/workload created by the simulator

Simulator

- L3 (GE) ISim
- PatrolSim, using ISim 4.0

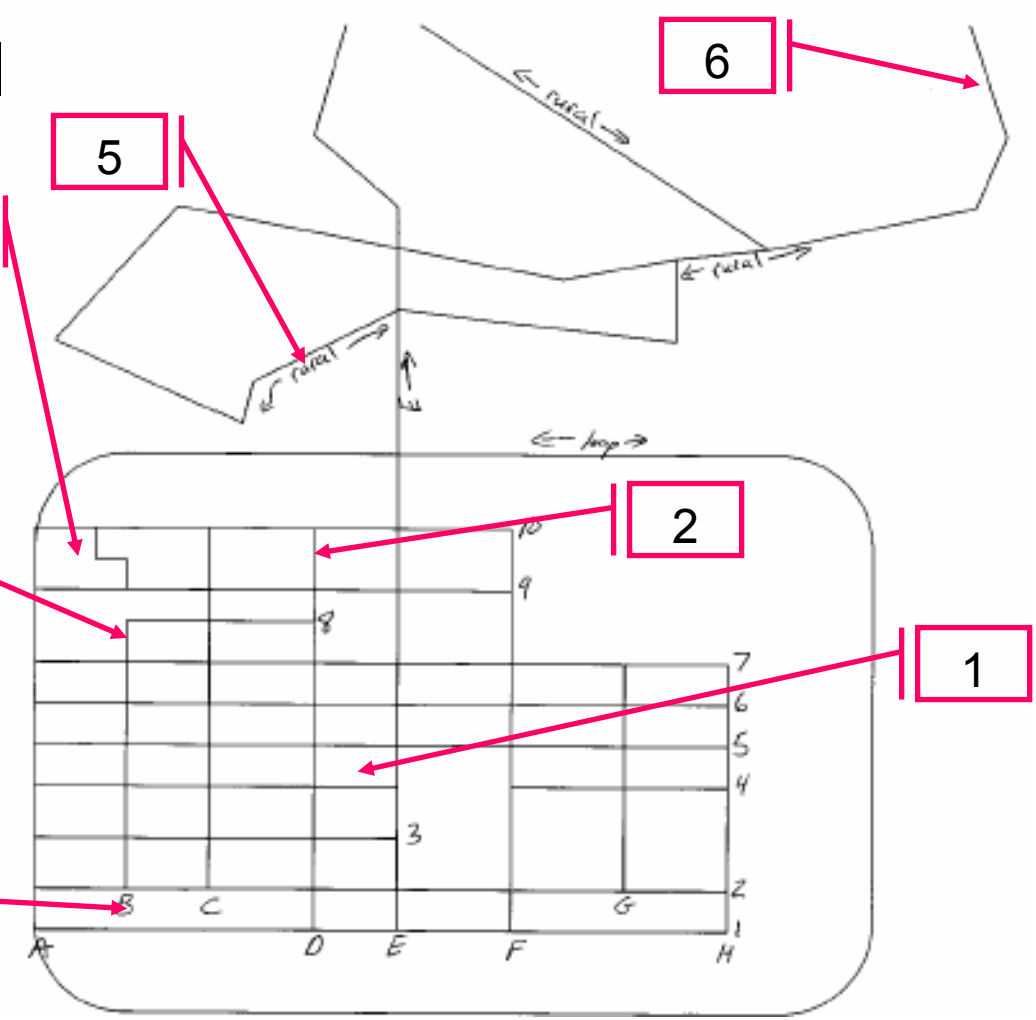




- 150° projection system
- Modified Crown Victoria cabin

The Virtual Environment

- Combined urban/rural environment
- Six markers to navigate between
- No traffic or pedestrians in the environment
- No weather or road abnormalities



Methodology

- Participants
  - Sex: 8 females, 5 males
  - Age: *M* = 22.78, *SD* = 4.51
  - All possessed valid driver's license
- NASA Task Load Index (TLX)
  - Hart & Staveland, 1988
- Simulator Sickness Questionnaire (SSQ)
  - Kennedy et al., 1993, RSK Assessments

Methodology

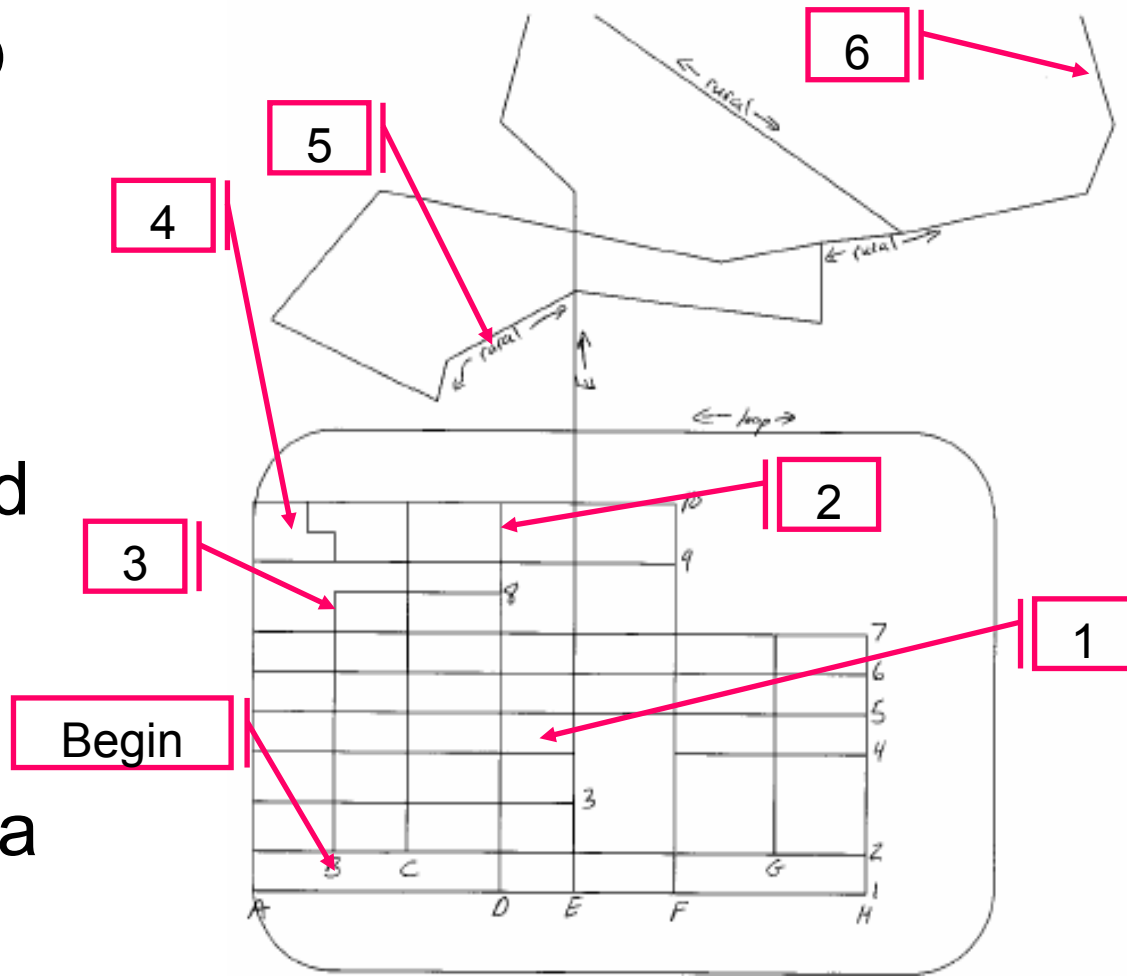
- Procedure
  - Three Phases
    - Introductory
    - Driving with a map and directions
    - Driving without a map and directions

Methodology

- Introductory Drive
  - Participant instructed to drive normally while following the speed limit
  - Given no objective or tasks
  - Phase duration: 5:00 min

Methodology

- Driving with a map and directions
  - Participant instructed to drive normally while following the speed limit
  - Task: using map and directions, navigate between a series of markers

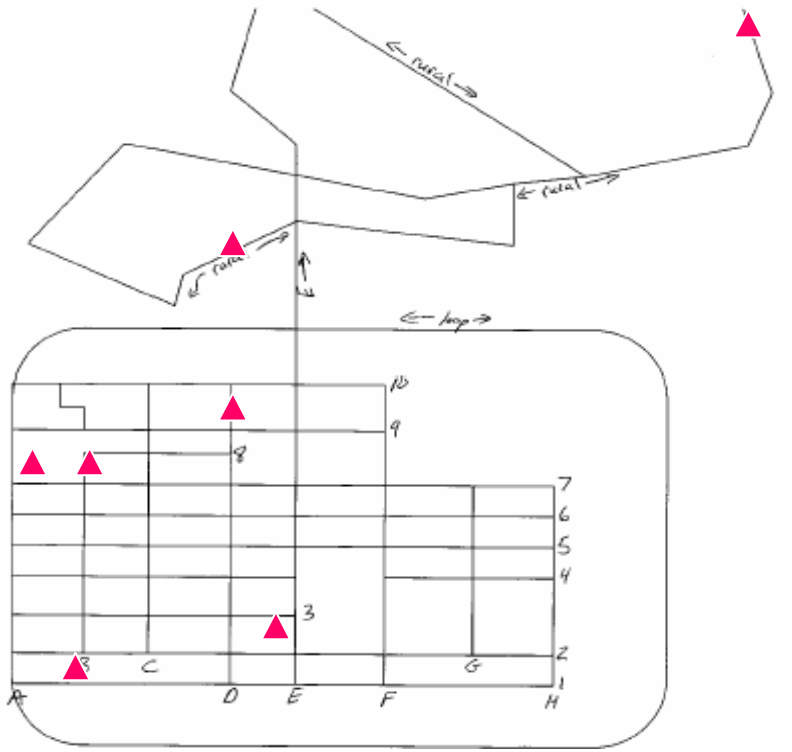


Methodology

- Driving without a map and directions
  - Participant instructed to drive normally while following the speed limit
  - Task: navigate between a series of markers, in reverse order, without aid of map or directions

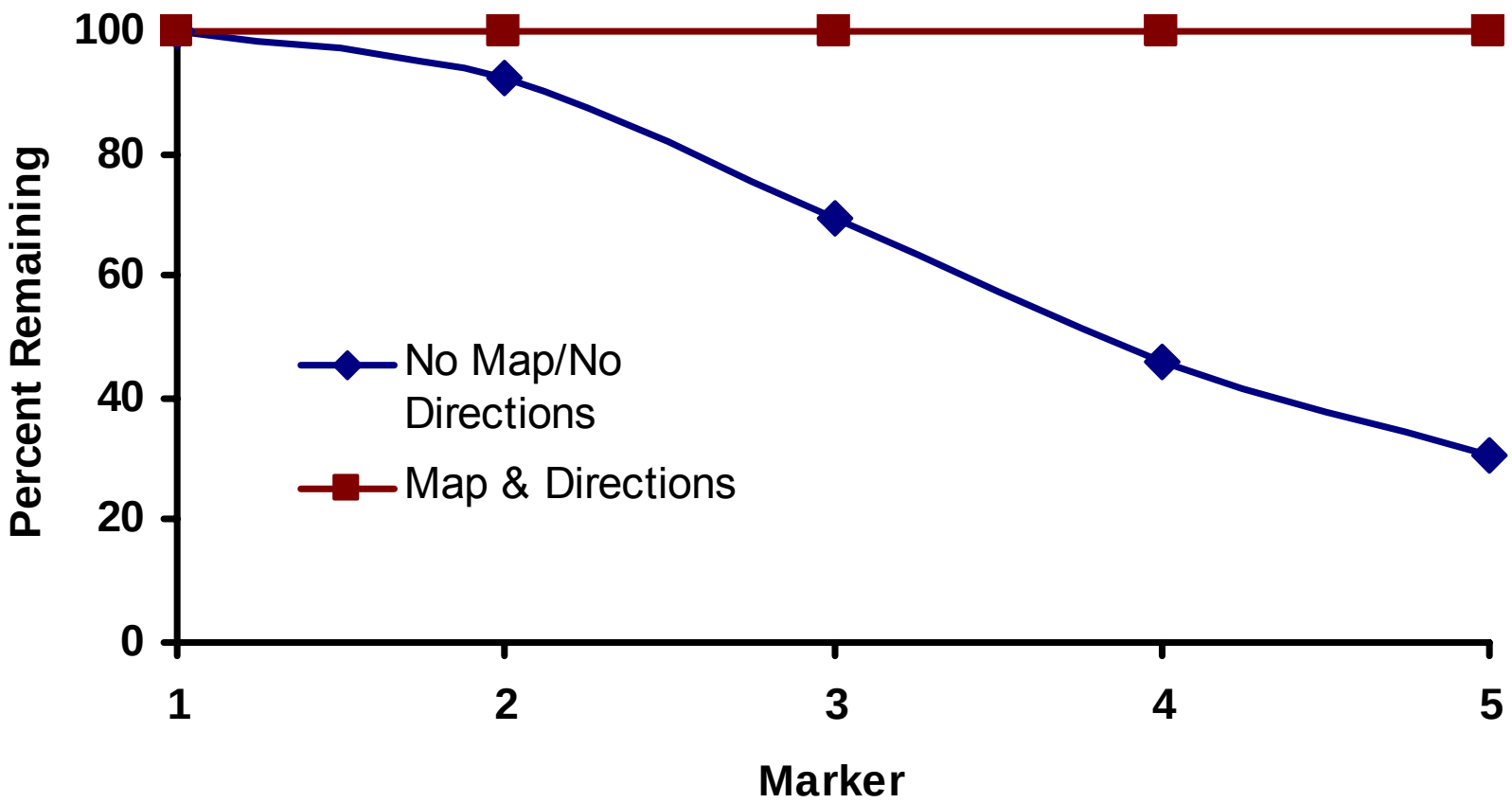
Equidistance of Phases

- Phases designed to be equidistant in both time and distance
- Time to complete phase data from the map & directions condition supports this

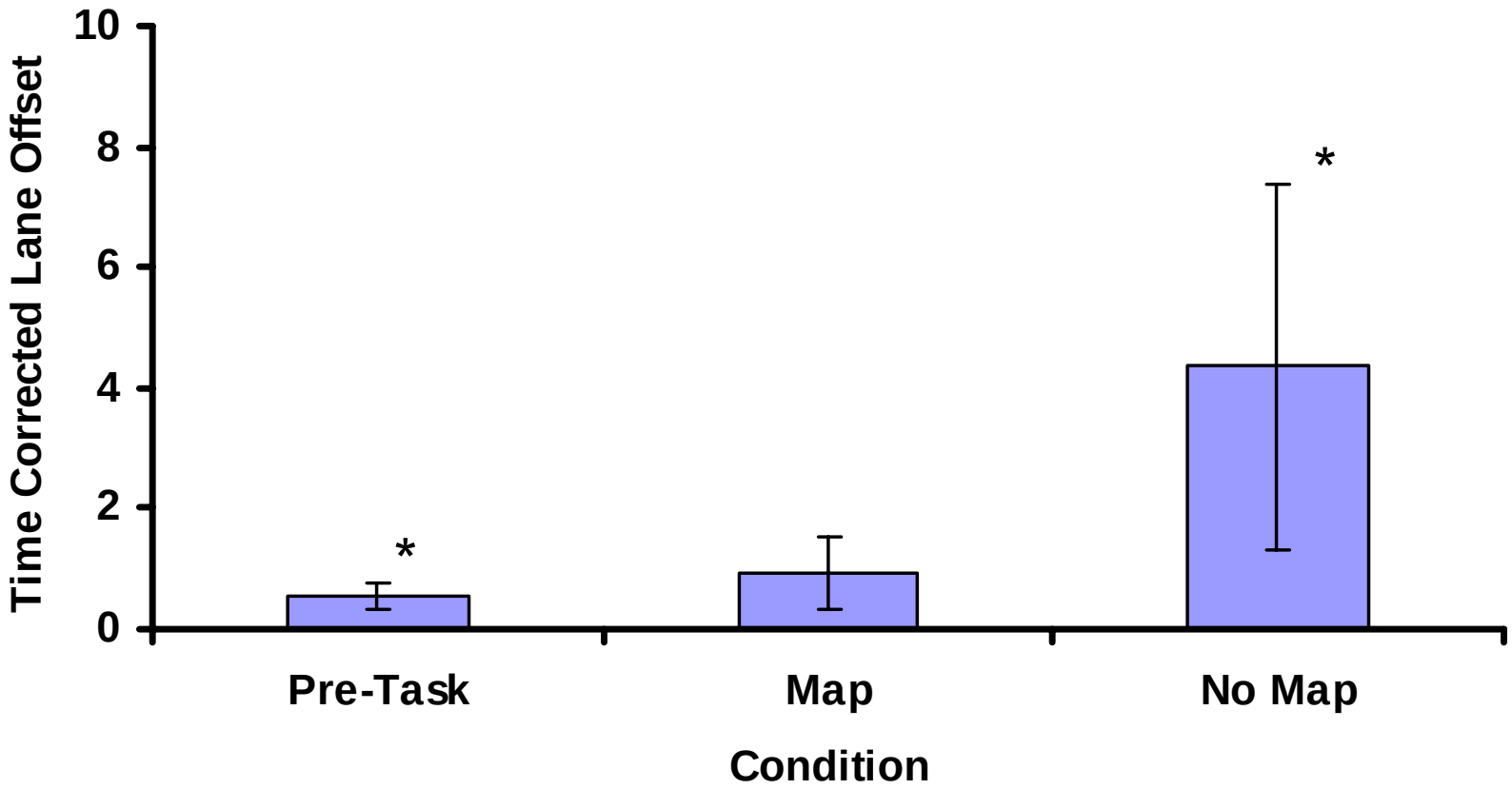


$F(4,48) = 1.108$, $p = .364$, part $\eta^2 = .085$

Completion of the No Map Condition

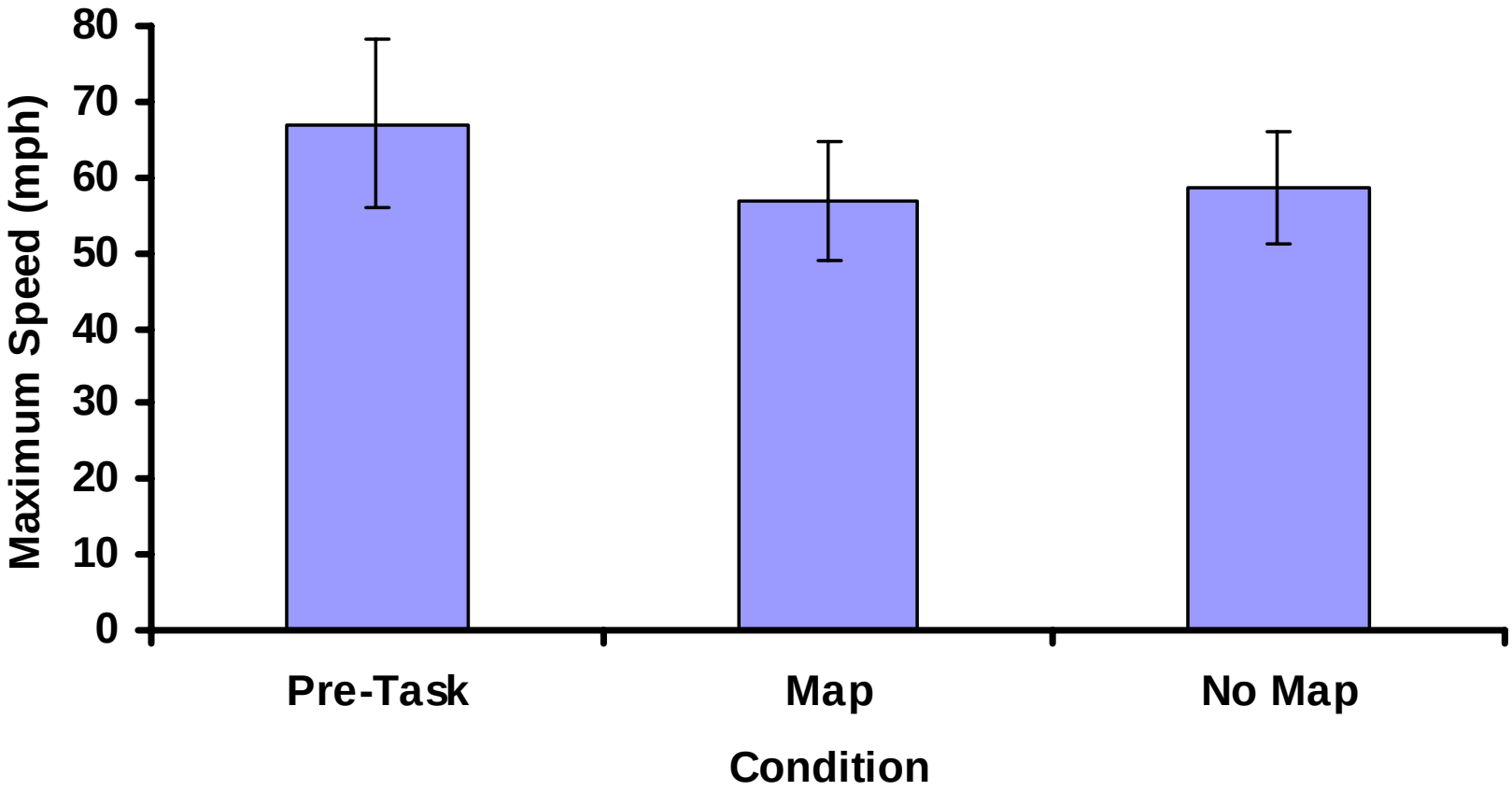


Lane Offset



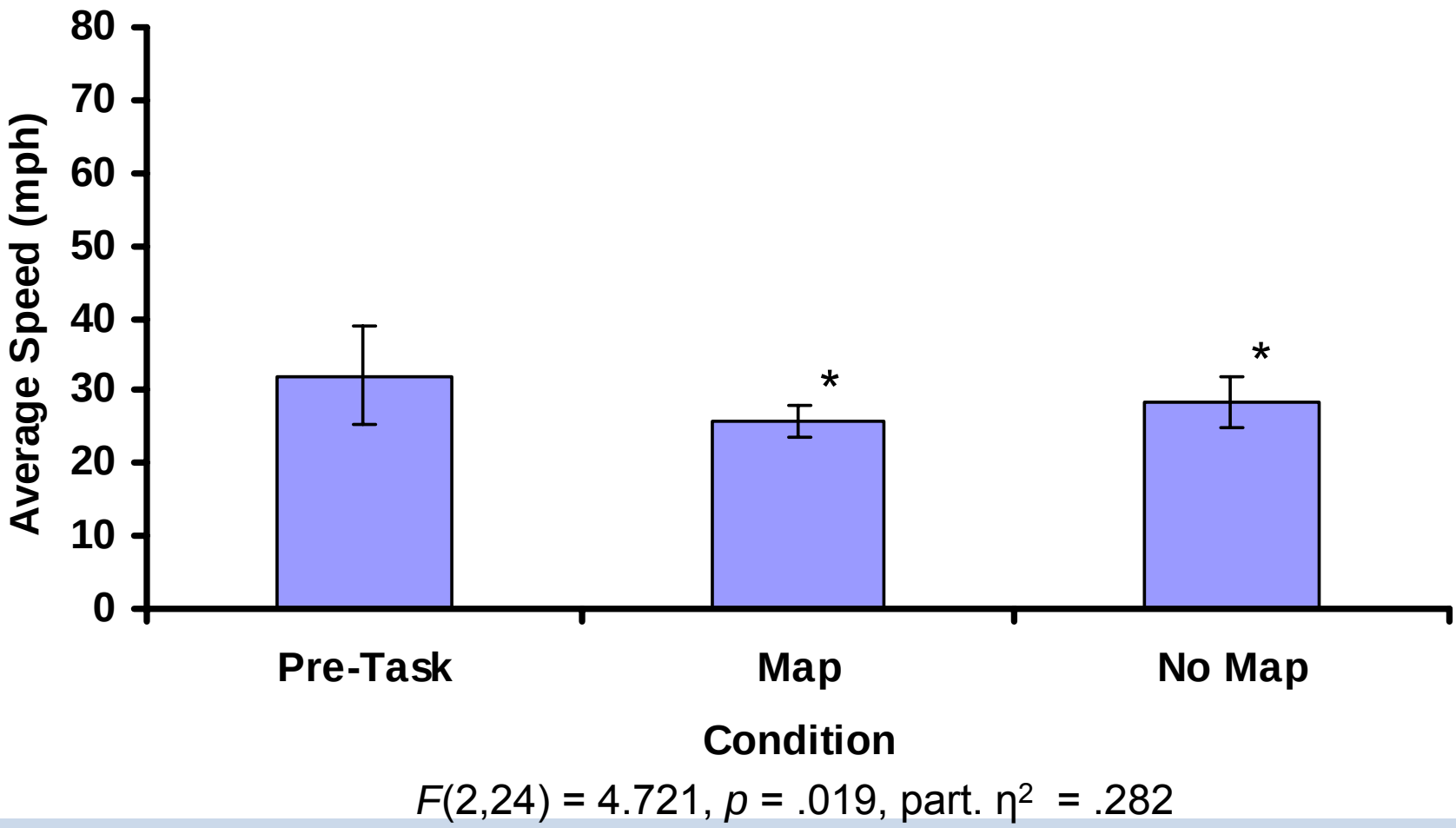
$F(2,24) = 7.578$, $p = .003$, part. $\eta^2 = .387$

Maximum Speed



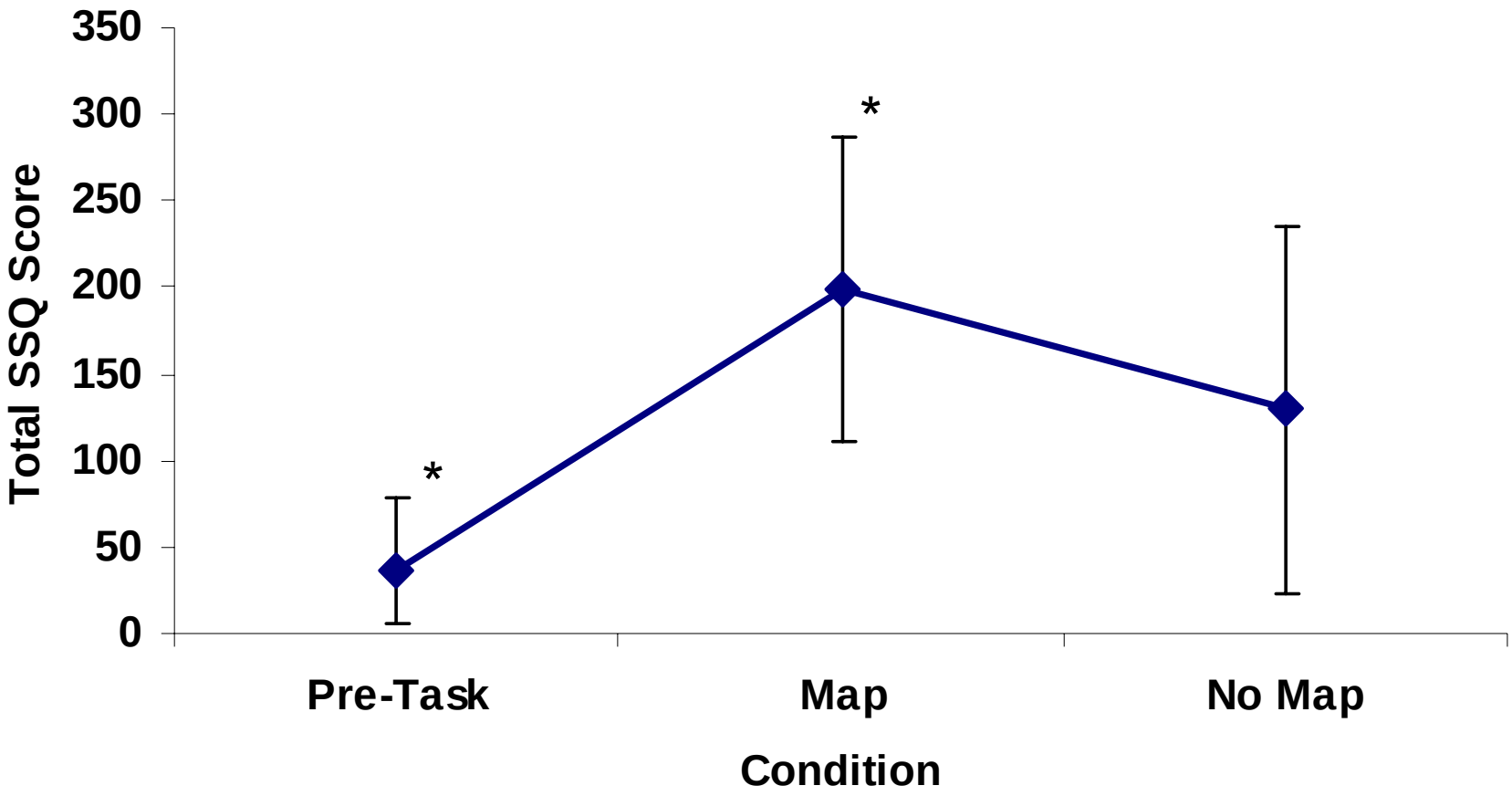
$F(2,24) = 3.313$, $p = .054$, part. $\eta^2 = .216$

Average Speed



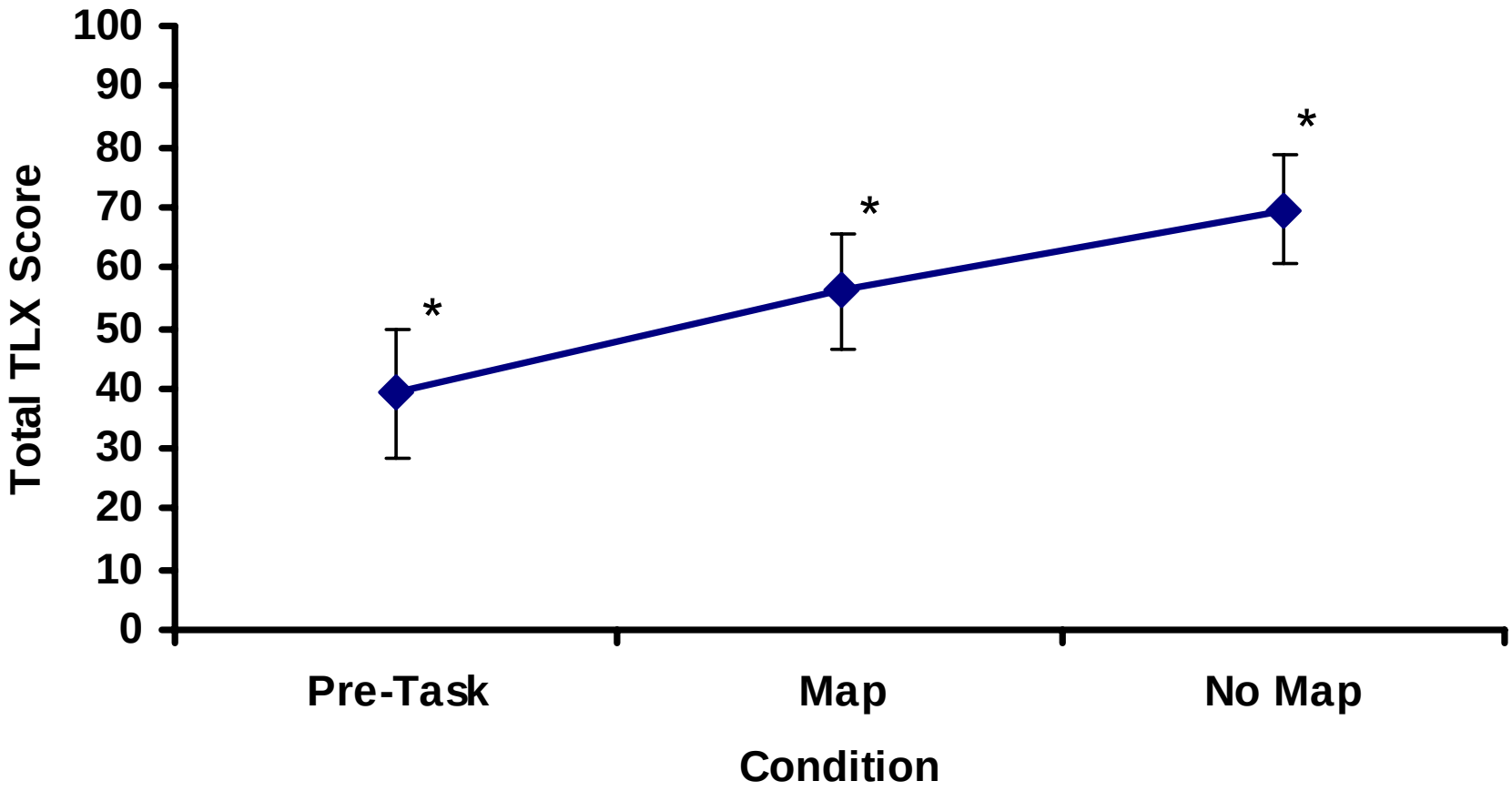
$F(2,24) = 4.721$, $p = .019$, part. $\eta^2 = .282$

Simulator Sickness Symptoms



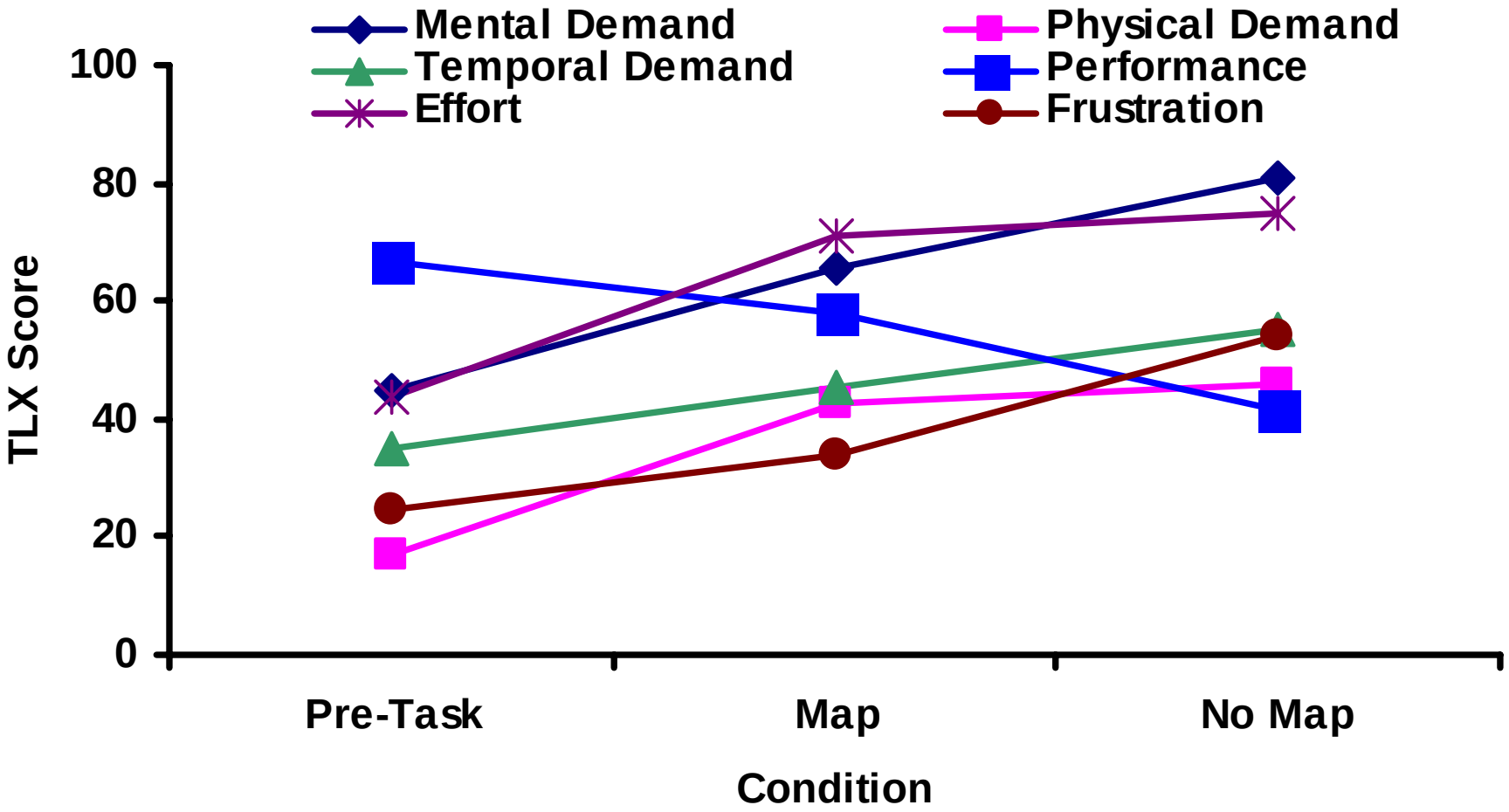
$F(2,24) = 4.941$, $p = .016$, part. $\eta 2 = .292$

Global TLX

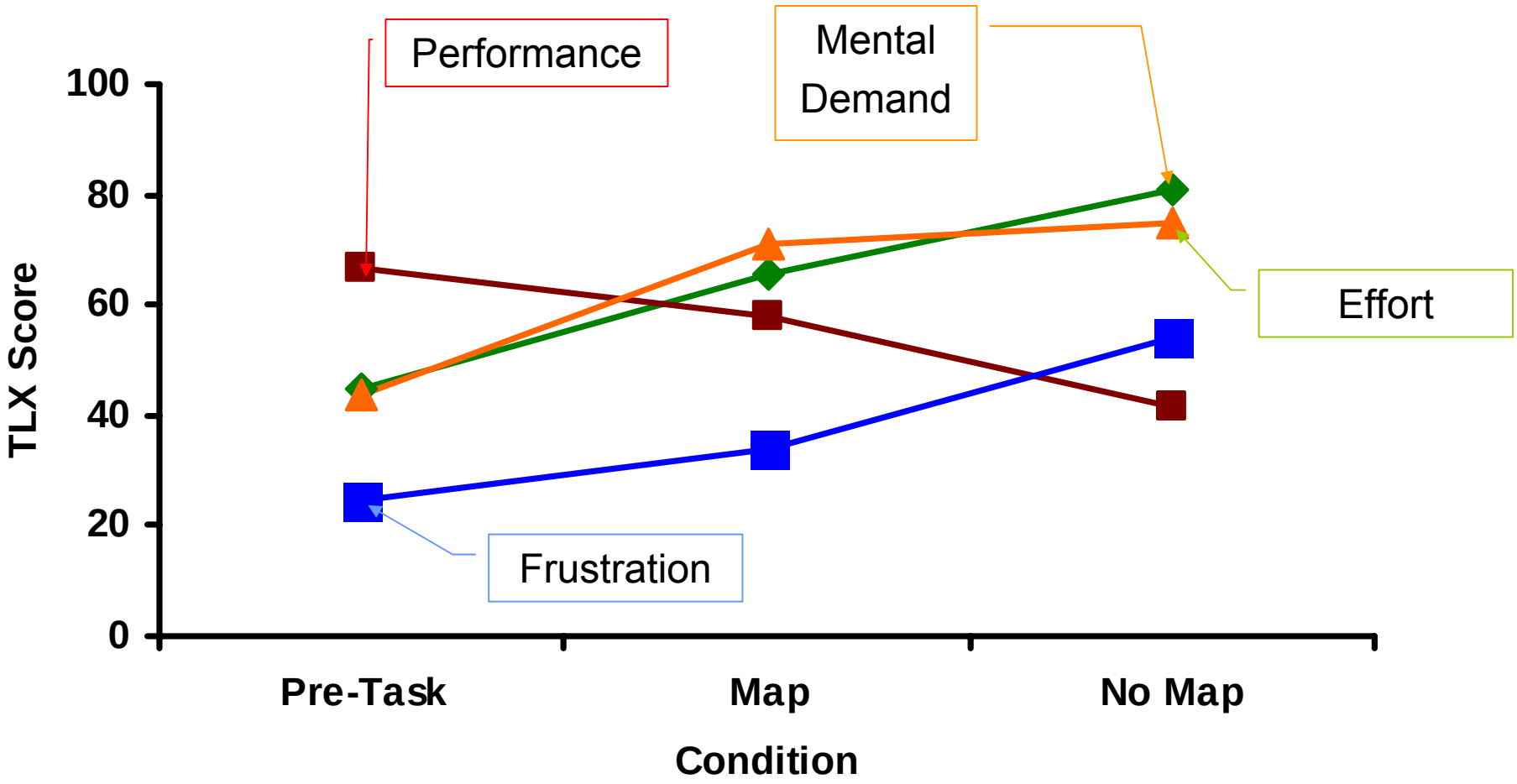


$F(2,24) = 29.346, p < .0005$, part. $\eta 2 = .710$

TLX Subscales



TLX Subscales



Conclusion

- Simulator Sickness Questionnaire scores stabilize
- Physical Demand (NASA TLX) stabilize
- Other workload indicators rise
- ∴ Measuring task induced workload

  Hysteresis effects

Implications

- Simulation is a research tool
- Further driving simulation studies
  - Decision Making
  - Friend/Foe Identification
- Other types of errors
  - Navigational
  - Unmanned Vehicles
  - Surgery







Acknowledgments

This work was facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program, P.A. Hancock, *Principal Investigator*, administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

This work was also facilitated by the FDOT, Tal Oron-Gilad, *Principal Investigator*, Sastry Putch, Ph.D Program Manager.

The driving simulator as a workload research tool.

Justin F. Morgan[1], Tal Oron-Gilad[1,2], and Peter A. Hancock[1,2]

1. The University of Central Florida Department of Psychology,
2. Institute for Simulation and Training

Introduction

The task of driving is inherently stressful. There are large demands made of the driver's cognitive abilities, as well as physiological demands that controlling the vehicle require. Driving is generally considered a high workload task, as evidenced by the large amount of research performed examining the mental workload aspects of driving (for a review see Recarte & Nunes, 2003).

Increasing mental workload on the part of the driver is associated with longer reaction times to signals such as changing traffic lights and oncoming traffic (Patten, Kircher, Östlund, & Nilsson, 2004). In fact, many types of in-vehicle displays have been shown to lower overall driver performance (Lansdown, Brook-Carter, & Kersloot, 2004). In general, increasing drivers' mental workload has lead to a reduction in objective measures of performance.

Although there is often disagreement about the definition and boundaries of the topic, workload is one of the most commonly studied topics in human performance. Over the decades, many assessment techniques have been developed for measuring human workload. Common objective and subjective measures of workload take into account factors such as mental workload, temporal demands, and effort in a particular task to generate an overall description of workload. These measures have proven to represent the workload experienced by the operator in a variety of settings and conditions.

Although workload research has occurred in a variety of settings, little research has been performed which uses the concept of driving, and in particular driving simulation, as a platform for workload research. The present study seeks to provide data which supports the use of a driving simulator for workload research by demonstrating that reliable and accurate measures from a well-regarded workload assessment instrument are available in a simulated driving environment and that changes in workload observed during the experiment were not due to physiological changes due to discomfort from the simulator.

Method

Participants

Eight adults (four males and four female) participated in this experiment. The five participants had an average age of 24.4 years (*SD* = 4.4). No particular inclusion/exclusion criteria were set for participation besides holding a valid driver's license, having normal color vision, and normal or corrected-to-normal vision. Participants were compensated for their participation with a payment voucher which can be exchanged for $7.50.

Simulator

One PatrolSim Mark III driving simulator was used for this experiment. The PatrolSim is a fixed-base, medium fidelity, driving simulator manufactured by L3. The simulator has all standard cabin controls of an automobile (steering, brake, throttle, turn signals, etc.), and sits in front of a projection screen which yields an area of approximately 150° horizontally which the participants were seated approximately three feet from.

Workload Measure

The NASA Task Load Index, or TLX, (Hart & Staveland, 1988) was administered via computer to each participant. The TLX is a multidimensional scale which measures six aspects of perceived workload (mental demand, physical demand, temporal demand, performance, effort, and frustration) on a scale from 0 to 100. In addition to the six individual scores and their weights, an overall value of perceived workload is generated. For the purpose of this research, only the combined, total workload score is examined.

Simulation Sickness Measure

Developed by Kennedy, et al. (1993), the Simulator Sickness Questionnaire (SSQ) is a widely used multidimensional measure of simulation sickness. The SSQ uses a pre-immersion interview and symptom checklist as a baseline to which subsequent symptom checklists are compared to. The SSQ measures symptoms clustered into three dimensions: nausea, oculomotor, and disorientation. For the purposes of this analysis, only the combined, total score is examined.

Procedure

After obtaining informed consent, the participant was administered the SSQ pre-immersion interview and symptom checklist forms. Following completion of these, the participant completed a baseline measure for the NASA TLX. Participants were then asked to enter the simulator. Each participant was first given a five minute period to familiarize them with the operation of the simulator. This initial exposure to the simulator was done in order to reduce the likelihood of the participant experiencing symptoms of simulator sickness and to reduce the participant's uncertainty with the control of the simulator.

Participants were then given a map and a set of turn-by-turn driving directions. Participants were instructed to proceed according to the directions to a series of six markers placed at approximately equal distances in the environment. When they reached each marker, participants were instructed to come to a complete stop, honk the automobile's horn, and then continue to the next marker. Participants were informed that they could refer to the map and driving directions at any point in the task. After completing the driving route, post task versions of the SSQ and NASA TLX were administered.

For the second driving task, participants began from the last marker and were asked to navigate through the markers in reverse order, without the aid of a map or

driving directions. Participants were instructed to continue as far as possible and that when they felt they were lost or unable to find the next marker, inform the experimenter of this and the experiment would stop. At completion of the route or after the participant could not find the next point, post task versions of the SSQ and NASA TLX were administered and the participant was debriefed.

Results

Completion of the task

All participants completed the entire route in the map condition. The number of participants who completed the no map condition declined as the task continued.

	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Participants Remaining	8	7	6	4	2
Percent of Participants Remaining	100	87.5	75	50	25

Table 1: Participant completion of the no map condition

When examined graphically, the percentage of participants completing the no map condition demonstrated a sharp decline (see figure 1).

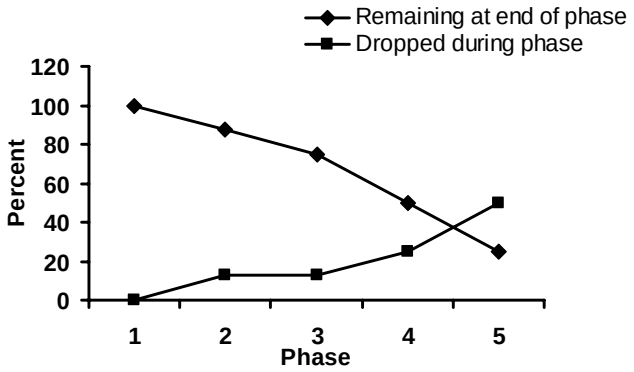


Figure 1: Percentage of participants dropped and remaining during no-map phase.

Comment [hf1]: Percentage is from 0-100 here you have probabilities – so you need to either change the graph axis title or multiply by 100

NASA TLX

A significant change was present in the workload scores across the three phases (practice session, driving with map, driving without map), $F(2,21) = 11.358$, $p < .0005$, part. $\eta^2 = .520$. Planned comparison analysis shows that the map ($M = 55.83$, $SD = 13.79$) and no map ($M = 72.00$, $SD = 4.49$) phases differed significantly from one another, while the baseline ($M = 38.50$, $SD = 19.56$) differed significantly only from the no map phase.

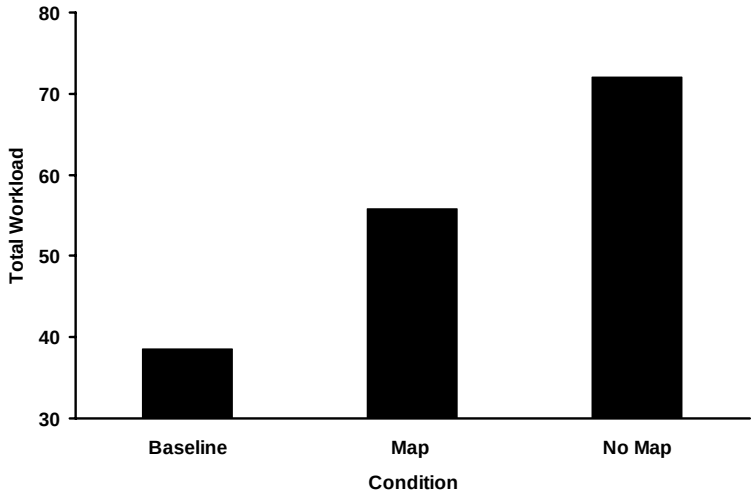


Figure 2: NASA TLX Scores by phase of participant drive

SSQ

There were no significant effects were present for the data from the SSQ, $F(2,21) = 2.264$, $p = .129$, part. $\eta^2 = .177$, an increase in SSQ total scores was observed between the baseline measure ($M = 57.49$, $SD = 84.89$), the first post-immersion measure after the map condition ($M = 179.11$, $SD = 133.24$) and the last measurement after the no map condition ($M = 199.21$, $SD = 193.29$).

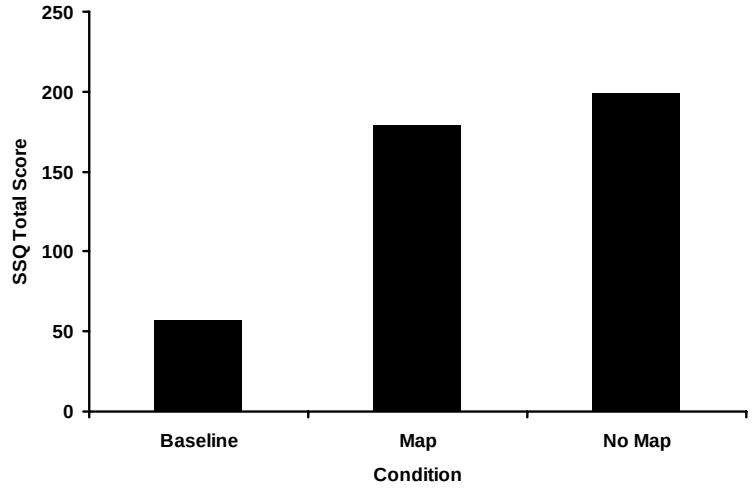


Figure 3: SSQ Scores by phase of participant drive

Discussion

The lack of any significant differences between time measures for driving different segments of the route demonstrated that the route was symmetrical and equidistant in layout. This allowed for the examination of the workload measures as a measure of the navigational task instead of a measure of the driving task. The lack of a significant difference between baseline and map driving conditions on the workload measure show that the actual task of driving, when no additional stressors are introduced, is not a significant contributor to workload. The addition of an additional workload item, in this case the navigation without a map, greatly increases workload over both baseline and map conditions. These support the hypothesis that the navigational task was the significant contributor to workload in this experiment.

Participants, on average, did not experience many symptoms of simulator sickness during the experiment. This is supported by the gradual, yet non-statistically significant, increase in scores on the SSQ. The SSQ displays a wide range of possible scores, and is extremely sensitive to changes in participant symptoms. This led to wide variability in scores, which likely contributed to the non-significance of the differences in conditions. However, the lack of a significant increase in the SSQ leads to the conclusion that workload was not being reliably influenced by any symptoms of simulator sickness, and was instead influenced solely by the navigational task.

Of interest in this research project is the validation of a ready to use driving simulator for workload research. We attempted to demonstrate this by showing that while driving an equidistant and symmetrical route under varying degrees of workload, high quality measures, which were due to the task characteristics, could be obtained and that the task's workload could be reliably separated from the workload imposed by driving. This was supported by the experimental findings, and supports the idea that a driving simulator can be used as a platform for workload research.

References

Hart, S.G. & Staveland, L.E. (1988). Development of the NASA TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. International Journal of Aviation Psychology, 3, 203-220.

Lansdown, T.C., Brook-Carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. Ergonomics, 47, 91-104.

Patten, C.J.D., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: Cognitive workload and attention resource allocation. Accident Analysis & Prevention, 36, 341-350.

Recarte, M.A., & Nunes, L.M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. Journal of Experimental Psychology: Applied, 9, 119-137.

APPLICATION OF FUZZY SIGNAL DETECTION THEORY TO THE DISCRIMINATION OF MORPHED TANK IMAGES

J.L. Szalma[1], T. Oron-Gilad[2], B. Saxton[1], and P.A. Hancock[1]

[1]*University of Central Florida*
Orlando, Florida

[2]*Ben-Gurion University*
Beer-Sheva, Israel

The effect of response set size on performance on a detection task was evaluated using both fuzzy and traditional signal detection theory. Fuzzy categories of stimuli were created using morphing software to blend profile images of American (M1A1) and Iraqi (T55) tanks to different degrees. These combinations were used to create static images varying from 100% T55 to 0% T55 (100% MIA1). Participants were asked to indicate the degree to which each image did not resemble an American tank. Consistent with previous research, results indicated that the FSDT model conforms to the normality assumption of traditional SDT. In addition, forcing observers to make binary decisions impaired performance relative to multi-category response sets in the FSDT analysis but not the traditional analysis. However, there were more model convergence failures in the FSDT analysis relative to the traditional analysis, mostly associated with conditions in which there were 100 response categories.

INTRODUCTION

Combat identification refers to the means by which military units distinguish friend from foe during operation. Accuracy in combat identification improves military effectiveness in two ways. First, it enhances the ability to successfully engage the enemy, but just as important it minimizes the risk of fratricide, which is the accidental destruction of friendly and or allied forces. Fratricide is a ubiquitous risk in warfare, accounting for up to 10 to 15% of friendly causalities during operations. In land-land operations some estimates have suggested figures closer to 40% (Bourn, 2002). In the first Persian Gulf conflict in Iraq, for example American forces accidentally destroyed nearly 80% (27 of 35) of the U.S. M1 Abrams and Bradley Fighting Vehicles lost in combat (Briggs & Goldberg, 1995). The process of attaining an accurate characterization for detected objects in combat environments with high confidence is a very difficult one, requiring timely application of tactical options.

Signal Detection Theory (SDT) is an extremely useful and powerful analytical tool for evaluating human and machine performance in situations which require detection and/or discrimination. However, traditional SDT artificially forces events into crisp, mutually exclusive categories (e.g., either black or white, you are either with us or against us), which may not be accurate representations of the true states of the world. In many instances events possess properties of both signal and non-signal, at least to some varying degree. Indeed, it is such blending that leads to uncertainty in decision making in many operational settings. Parasuraman, Masalonis, and Hancock (2000; see also Hancock, Masalonis, & Parasuraman, 2000) have addressed this uncertainty by combining elements of SDT with those of Fuzzy Set Theory, in which category membership is not considered mutually exclusive and stimuli and responses to them can therefore be simultaneously assigned to more than one category. Thus, in Fuzzy SDT (FSDT) a given stimulus may be categorized as both a correct detection and a false alarm depending on the relative degrees of signal-like versus non-signal-like properties. Formally, in FSDT a stimulus dimension, **s**, and a response dimension, **r**, are defined from 0 to 1, with 0 representing full membership in the category 'non-signal' (or 'no' response) and 1 representing full membership in the category 'signal' (or 'yes' response). For instance, a value of **s**=.5 indicates that the stimulus possesses characteristics of signal and non-signal to equal degrees, and **r**=.5 indicates maximal uncertainty in responding (e.g., "not sure"). Each dimension is related to real world stimulus and response properties via a mapping function; the key to successful application of FSDT. For instance, a fuzzy signal of **s**=.5, a point of maximal uncertainty along the fuzzy dimension, must be defined by a particular characteristic or set of characteristics of the real-world variable (e.g., the percentage of a morphed image that is derived from a T55 Iraqi tank; for more detail on mapping functions, see Parasuraman et al., 2000).

Initial investigation comparing FSDT and standard SDT indicated that FSDT analysis generally reduces the false alarm rate and inflates the hit rate (Masalonis & Parasuraman, 2003; Murphy, Szalma, & Hancock, 2004). In addition, there is evidence that FSDT meets the Gaussian assumptions of traditional SDT and that fuzzy ROC functions are of the same general form as traditional or crisp ROCs (Murphy et al., 2004). The current study was designed to determine if these ROC results extend to more complex stimuli that simulate real-world objects, and to examine the effect of response set size on fuzzy ROCs. Thus, a 5-category condition was compared to a 100 category condition and a binary response set. Although the latter is not truly a fuzzy response, Parasuraman et al. (2000) noted that there is no constraint that both the **s** and **r** dimensions be fuzzy. Indeed, examination of binary response to fuzzy stimuli is critical for evaluating

performance for domains in which discrete yes/no decisions are required.

EXPERIMENTAL METHOD

Three levels of response bias instruction set (conservative, unbiased, and lenient) were factorially combined with three response sets (2, 5, or 100 categories) to yield 9 experimental conditions. Response bias was manipulated using a point system as a payoff matrix (see Macmillan & Creelman, 2005). All participants experienced all conditions. The order of conditions was counterbalanced separately for each factor across participants, with the constraint that in any given session observers experienced only one bias condition. This was done to avoid problems associated with shifting criteria within a session. Prior to these experimental conditions participants engaged in a practice session in which they experienced each response set without bias instructions. Six students from the University of Central Florida (3 men and 3 women) participated in this study for monetary compensation.

The stimuli for this experiment consisted of profile images of an M1A1 (US) tank and a T55 (Iraqi) tank. These images were combined using morphing software, so that as the two images are blended, they assume characteristics of both tanks. Thirty blended images were selected for presentation, ranging from 38% T55 to 60% T55. The selection of this range was based on work in which the psychophysical function of these stimuli was assessed using the method of limits (Oron-Gilad, Morgan & Hancock, under revision). Images beyond 60% T55 were almost always labeled as 'Iraqi' and images below 38% were almost always labeled 'American.' Examples of the two extremes and a morphing of 50% are shown in Figure 1.

In each condition observers were presented with 480 static images, with each of the 30 images presented 16 times. On each trial observers were asked to indicate the degree to which the image appeared to be an Iraqi tank.

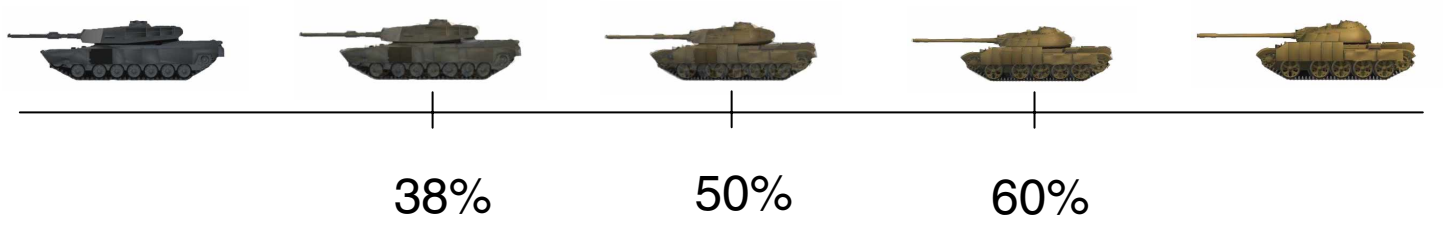


Figure 1. *Examples of the stimuli used in this experiment. The tank on the far left is an American M1A1, and the tank on the right is an Iraqi T55. The middle image is the 50/50 morphed combination of the two tanks.*

RESULTS

Current algorithms for estimation of ROC curves require provision of frequencies of response. However, these programs do not permit application of fractional frequencies such as those associated with an FSDT analysis. Due to this constraint ROC functions were estimated by rounding the frequencies to the nearest whole number and entering them into FitRoc: Parameter Estimation for Gaussian Signal Detection Model (Wickens, 2002), version 2.1 program which provides (1) a test of how well an ROC curve fits the Gaussian model (χ^2), (2) an estimate of perceptual sensitivity (A_z), and (3) an estimate of response bias (β_{log}). Additionally, the intercept and slope for the z-score form of the ROC are estimated by the program ($z_H = a + b\ z_F$). Note that methods for determining SDT parameters are the same for both fuzzy and traditional analyses. The two models differ in how hit and false alarm rates are calculated. The formulas for calculating FSDT measures were obtained from Parasuraman et al. (2000), and the method for computing crisp hit and false alarm rates followed procedures outlined by MacMillan and Creelman (2005). For the latter analysis the stimulus and response dimensions were collapsed into two categories in order to facilitate direct comparison of the traditional and fuzzy SDT models. We also compared response set conditions using a z test as described in Wickens (2002).

Tables 1 and 2 report goodness of fit, intercept (i.e., 'a'), slope (i.e., 'b'), and perceptual sensitivity estimates for the fuzzy and traditional SDT methods, respectively.

Fuzzy SDT Analysis

Binary Response Set. Without exception, the data for every observer met the Gaussian equal variance assumptions of SDT.

5 Category Response Set. The equal variance Gaussian model fit for observers 2, 3, and 6. The unequal variance model fit for observer 5. Neither model fit for observers 1 and 4 (see Table 1). Interestingly, observer 1 only used the two extreme responses (1 and 5) for the lenient and conservative conditions. Although observer 5 did use other categories, the distribution of responses was bimodal, with the majority of responses at the extremes. Thus, the failure of the data of these two individuals to meet the equal variance assumption may be in part due to response strategy.

100 Category Response Set. The data of two observers, 2 and 3, met the equal variance assumption. The unequal variance model fit the data of observer 6. However, the unequal variance model produced an extremely steep slope, a

very unusual outcome, given that slopes generally vary from .5 to 2 (Swets, 1996). Neither model fit for observers 1, 4, and 5. However, observers 1 and 5, primarily used extreme categories in responding, as in the 5-category response condition.

Comparison of Response Sets: Fuzzy SDT Analysis

2 category vs. 5 category response. For those observers whose data fit the Gaussian SDT model, the ROCs were compared for these two conditions. In the case of observer 5, whose data met the unequal variance assumption, the unequal variance model for the 2 category condition was used for comparison. In each case the ROCs for the two conditions were significantly different, with the 5-category response set producing the higher sensitivity. Thus, constraining response sets impairs detection performance.

2 category vs. 100 category response. For the three observers for whom comparisons could be made, in each case the ROCs were significantly different, with the 100 category response set yielding higher sensitivity.

5 category vs. 100 category response. For the three observers for whom comparisons could be made, a significant difference in ROCs was observed for observer 2, with the 100 category condition yielding the higher sensitivity. In the other two cases, the two conditions did not differ in sensitivity. Indeed, for observer 3 the points were clustered close together in ROC space.

In sum, binary response sets seem to constrain performance as evaluated by FSDT. The number of multiple categories (5 vs. 100) exerts a weaker effect, showing differences for only one person. For that individual, providing more response categories improves detection, presumably because it allows a more precise linkage of response to the degree of signalness observed. However, three observes chose the response strategy of using primarily extreme categories and therefore did not utilize the full range of possible responses.

Table 1. *Displays goodness of fit, sensitivity, and criterion for bias statistics calculated using Fuzzy SDT*

Participant	χ^2	A(z)	a	b	Conservative β_{log}	Unbiased β_{log}	Lenient β_{log}
Binary Response							
1	.256[e]	.787	1.126	1.00	.037	-.007	-.212
2	.177[e]	.790	1.139	1.00	.749	.302	.106
3	.020[e]	.793	1.157	1.00	.023	-.075	-.186
4	.208[e]	.785	1.117	1.00	-.073	-.204	-.656
5	.398[e]	.780	1.091	1.00	-.076	-.272	-.690
6	2.517[e]	.746	.938	1.00	-.238	-.185	-.587
5 Category							
1	cf						
2	.964[e]	.910	1.895	1.00	3.616	-.640	1.826
3	.449[e]	.940	2.201	1.00	-.394	-.131	-.454
4	cf						
5	.126[u]	.923	1.426	.053	-3.485	-3.740	-3.932
6	2.823[e]	.917	1.954	1.00	-.231	-.103	-1.445
100 Category							
1	cf						
2	.734[e]	.942	2.224	1.00	3.203	.607	.941
3	1.045[e]	.950	2.329	1.00	-.464	-.104	-.698
4	cf						
5	cf						
6	.224[u]	.860	15.023	13.896	2.255	2.668	1.872

[u] data fits the unequal variance model; [e] data fits the equal variance model; cf. = convergence failure.

Traditional (Crisp) SDT Analysis

Arguably the most direct comparison of a traditional SDT model to the FSDT analysis described above is to reduce the number of stimulus and response categories to two. Thus, an analysis was computed in which a response of 3 or lower was categorized as a 'no,' and a similar analysis in which a response of 3 or higher was considered a 'yes.' These analyses differ only in whether the middle category is categorized as 'yes' or 'no.'

Binary Response Set. The Gaussian assumption was met for the data of each participant, and with the exception of participant 6, the equal variance assumption also held.

5-Category Response Set: Middle category as 'yes.' The Gaussian equal variance model fit for observers 1, 2, 4, and 5. The unequal variance model fit for observer 3, and neither model fit for observer 6.

5-Category Response Set: Middle category as 'no.' The Gaussian equal variance model fit for all participants except observer 6, for whom the unequal variance model fit.

100-Category Response Set. The Gaussian equal variance model fit for observers 1 and 4, and the unequal variance model fit for observers 2, 3, and 6. Neither model fit for observer 5.

Comparison of Response Sets: Traditional SDT Analysis

Pairwise comparisons of response set conditions were computed using the z test described in Wickens (2002). Comparisons were made only when the Gaussian model fit for each condition.

2 category vs. 5 category (middle category as 'yes'). Comparisons were possible for all observers except number 6. Results were mixed. For observers 1 and 3 there were no significant differences between these conditions. For observer 2 the binary response set yielded significantly greater sensitivity than the 5 category set. However, the opposite was true for observer 4 (see Table 2).

2 category vs. 5 category (middle category as 'no). Results for this comparison were the same as those observed when the middle response category was categorized as 'yes,' with the exception that in this case the binary response set produced higher sensitivity than the 5-category set for observer 3 as well as for observer 2. Comparison for observer 6 indicated no significant differences between conditions.

2 category vs. 100 category response. Comparisons of these conditions resulted in only one significant difference. The sensitivity of observer 1 was significantly greater in the 100 category condition relative to the binary response set.

5 category vs. 100 category response. Regardless of whether the middle response category was coded as a 'yes' or a 'no,' z tests revealed that for observers 1 and 2 the 100 category conditions produced significantly higher sensitivities relative to the 5 category condition. Comparisons for the other participants indicated no significant differences between these conditions.

DISCUSSION

The fuzzy analysis yielded similar results for the 5- and 100-category response set conditions. In the crisp analysis a similar pattern was observed with the exception of two participants, for whom the larger response set was related to higher sensitivities. However, there were differences between multiple category response and binary response set conditions for both analyses. In the fuzzy analysis performance was degraded by forcing a binary response. By contrast, results were mixed in the crisp analysis, with multi category response sets producing greater sensitivities in some cases and the binary response set producing superior performance in other cases.

In a traditional analysis a response is a correct detection, correct rejection, or an error of omission or commission, while in an FSDT analysis observers are awarded a degree of correctness in responding. Permitting multiple response categories and submitting the data to FSDT analysis capture these aspects of performance. That is, partial hits in fuzzy SDT are categorized as full hits or full misses in the crisp analysis, and partial false alarms are categorized in the traditional SDT analyses as full false alarms or correct rejections. The forced assignment of stimuli and responses to crisp categories may account for the inconsistencies in comparisons between conditions. It may be that the loss of information that results from such procedures biases the results in one direction or the other depending on how the participant's responses are distributed. Recall that a few participants used only extreme response categories in the multiple response set conditions. The outcome of a crisp analysis may therefore be expected to depend in part on the response pattern of the individual regardless of that person's perceptual sensitivity.

A central question for fuzzy SDT is whether the assumptions of traditional SDT hold for this new model. The current study confirms prior findings that, in general, the assumption of normally distributed noise and signal plus noise distribution holds (see χ^2 tests in Tables 1 and 2), and that these results extend to tasks in which the stimuli are complex visual images.

Future investigations will examine the effect of different distributions of stimuli (e.g., more signal-like than non-signal-like stimuli are presented), analogous to manipulations of signal rate in traditional SDT. In addition, future experiments will manipulate instructions regarding the definition of 'signal' and 'non-signal' to determine their effect not only on sensitivity but also on criterion setting. Recent evidence indicates that the fuzzy nature of the stimulus dimension may make stable criterion setting difficult for observers, which could account for cases in which the equal variance assumption was not met (Szalma, Parasuraman, & Hancock, in preparation).

REFERENCES

Bourn, J. (2002). *Ministry of Defence: Combat Identification, British National Audit Office Report*, retrieved August 1st 2005 from www.nao.gov.uk.

Briggs, R.W. & Goldberg, J.H. (1995). Battlefield recognition of armored vehicles, *Human Factors*, *37*(3), 596-610.

Hancock, P.A., Masalonis, A.J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science*, *1*, 207-230.

Macmillan, N.A., & Creelman, C.D. (2005). *Detection theory: A user's guide, 2nd ed.* Mahwah, NJ: Erlbaum.

Masalonis, A.J., & Parasuraman, R. (2003). Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics*, *46*, 1045-1074.

Murphy, L., Szalma, J.L., & Hancock, P.A. (2004). Comparison of fuzzy signal detection and traditional signal detection theory: Analysis of duration discrimination of brief light flashes.

Proceedings of the Human Factors and Ergonomics Society, 48, 2494-2498.

Oron-Gilad, T. Morgan J. F., & Hancock P. A. (,under revision). *Categorical Perception of Morphed Objects*

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42*, 636-659.

Szalma, J.L., Parasuraman, R., & Hancock, P.A. *Fuzzy signal detection theory: The form of empirical fuzzy receiver operating characteristics*. Manuscript in preparation.

Swets, J.A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics*. Mahwah, NJ: Erlbaum.

Wickens, T.D. (2002). *Elementary signal detection theory*. Oxford University Press.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. P.A. Hancock*, Principal Investigator.* The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

Table 2. *Displays goodness of fit, sensitivity, and criterion for bias statistics calculated using the Traditional SDT.*

Participant	χ^2	A(z)	a	b	Conservative β_{log}	Unbiased β_{log}	Lenient β_{log}
Binary Response							
1	1.023[e]	.940	2.196	1.00	.113	-.022	-.635
2	1.696[e]	.964	2.540	1.00	2.779	1.221	.452
3	.563[e]	.959	2.461	1.00	.111	-.299	-.695
4	.548[e]	.946	2.273	1.00	-.243	-.667	-2.046
5	3.011[e]	.939	2.191	1.00	-.206	-.828	-2.019
6	.651[u]	.805	2.997	3.342	.570	.826	-1.632
5 Category 3 as 'yes'							
1	1.077[e]	.945	2.263	1.00	-.819	-.837	-.931
2	3.723[e]	.926	2.044	1.00	5.235	-1.226	2.267
3	.114[u]	.955	1.816	.389	-2.509	-2.091	-2.435
4	.683[e]	.975	2.783	1.00	-3.034	-2.883	-2.813
5	4.239[e]	.944	2.250	1.00	-1.729	-2.450	-2.217
6	cf						
5 Category 3 as 'no'							
1	1.077[e]	.945	2.263	1.00	-.819	-.837	-.931
2	3.721[e]	.921	2.000	1.00	5.231	-.900	2.244
3	.630[e]	.933	2.117	1.00	.922	.878	.605
4	.217[e]	.972	2.703	1.00	-.194	-.082	-2.478
5	2.372[e]	.947	2.285	1.00	.524	.075	-1.241
6	2.599[u]	.838	6.724	6.749	2.093	2.245	1.133
100 Category							
1	5.410[e] *	.969	2.634	1.00	-.478	-.187	-.620
2	.780[u]	.958	3.930	2.044	3.011	1.199	1.536
3	.005[u]	.878	3.998	3.288	-.652	.692	.229
4	2.425[e]	.964	2.542	1.00	-2.121	-2.125	-2.755
5	5.437[n]	.943	1.795	.539	-.291	-1.317	-2.065
6	1.709[u]	.863	6.757	6.093	.729	1.708	-3.215

[u] data fits the unequal variance model; [e] data fits the equal variance model; [n] neither model fit—values for unequal variance model are shown; cf. convergence failure; * model fit was marginal ($.05<p<.10$).
'3 as yes' = middle response category coded as 'yes.' '3 as no' = middle response category coded as a 'no.'

PERFORMANCE, WORKLOAD, AND STRESS IN VIGILANCE: THE POWER OF CHOICE

J.L. Szalma and P.A. Hancock

University of Central Florida
Orlando, Florida

Observers were provided the illusion of choice by permitting them to 'choose' an easy or a hard version of a vigilance task. Participants were then assigned at random to either the condition they expressed preference for or their non-preferred condition. Participants in the comparison control condition were not offered the opportunity to choose. Task demand was manipulated using event rate variation, divided into high and low event rates. Results indicated that permitting participants a choice regarding 'difficulty level' and subsequently assigning them to that level enhanced signal detections but did not reduce their perceived workload and stress. In contrast, offering a choice and then assigning observers to their non-preferred condition impaired performance relative to a condition in which no choice was provided. This pattern of effects was more pronounced at the low compared to the high event rate. These results confirm the importance of motivational effects in vigilance and the impact choice has on performance in tasks requiring sustained attention.

INTRODUCTION

Vigilance represents the capacity to maintain attention over prolonged periods of time. It is an essential component of monitoring tasks performed by operators of modern automated systems. One of the most reliable findings in vigilance is the decline in performance with time on watch or the vigilance decrement function (See, Warm, Howe, & Dember, 1995). As well as this systematic diminution in performance capacity over time, vigilance also imposes a considerable degree of workload and stress on operators (Hancock & Warm, 1989; Szalma, Hancock, Dember, & Warm, 2006; Szalma et al., 2004; Warm, Dember, & Hancock, 1996). However, the demands of modern automation and current security concerns render vigilance tasks vital to public safety and health (Hancock & Szalma, 2003). It is therefore crucial to understand the factors that affect vigilance and how potential decrements in such performance may be ameliorated. There is considerable evidence that the characteristics of the task itself are critical determinants of performance (e.g., see Davies & Parasuraman, 1982; Warm & Dember, 1998). One such characteristic is the limited autonomy of the individual operator engaged in these often tedious tasks with little or no control over how their own performance. Indeed, Scerbo (1998) and Hancock (1998) have argued that a considerable degree of the stress of vigilance may derive from this imposition of the task on the individual by an external authority that permits little autonomy for action. Evidence to date indicates that when observers are allowed some measure of perceived or real control over the task environment, performance is enhanced. Thus, Gunn et al. (2005) reported that when observers in a simulated UAV task were permitted to take action against targets they detected, their performance improved and stress symptoms decreased. Further, even the illusion of control can facilitate performance. Dember, Galinsky, and Warm (1992) asked observers whether they wished to experience an 'easy' or 'hard' version of a vigilance task, and then randomly assigned them to one of two signal salience conditions. They reported that the provision of 'choice' attenuated the performance decrement relative to observers in a control condition who were not permitted such a choice.

One component purpose of the present study then was to extend the results of Dember et al. (1992) by manipulating not only whether observers have a choice in which task they experience, but also whether they were actually assigned their preferred condition. In other words, among observers provided the illusion of choice, half experienced a condition matching their preference while the other half received the version of the task they did not prefer. Unlike Dember et al. (1992), who manipulated task difficulty by signal salience, in our experiment demand was manipulated using event rate. Event rate was chosen because it is a task characteristic that exerts substantial influence on performance, and as such tends to show substantial performance differences (Warm & Jerison, 1984). It is important to note that the participants were unaware of these different event rate conditions when they made their 'choice.' Finally, the study reported by Dember and his colleagues (1992) examined the influence of choice on quality of performance. The current study included evaluation of the effect of choice on perceived workload and stress as well as objective performance outcome.

We hypothesized that an attenuation of the vigilance decrement would be observed for individuals who received a 'choice,' and that such performance benefits would be greater for those whose assignment matched their preference. In addition, these effects were expected to occur at both low and high event rates, but would be more pronounced in the latter condition because of the associated higher level of task demand. Specifically, we hypothesized that the benefit of choice would be greater in the context of the more demanding version of the task. It was also anticipated that the performance benefits in the choice conditions would result in reduced workload and stress relative to a control condition in which participants received no opportunity for choice, and that again this benefit would be greater at the higher event rate relative to the low event rate condition.

EXPERIMENTAL METHOD

Seventy-two undergraduates at the University of Central Florida served as participants in this experiment. They were paid $7.50 per hour for such participation. All observers had normal or corrected-to-normal vision and were free of known hearing impairments.

The study design was a 3 (choice: matched, mismatched, and no-choice) by 2 (event rate: high, low) by 4 (period) mixed design with repeated measures on the last factor. Twelve participants were assigned at random to one of the six experimental groups, with the restriction that in the choice conditions an equal number of individuals in each event rate condition were assigned to the choice-matched and choice-mismatched conditions. Thus, of the 24 participants in the choice-matched group, 12 received the low event rate and an equal number the high-event rate.

Participants experienced a forty minute vigil divided into four continuous 10-minute periods. They were not informed of the length of the vigil except that it would not exceed 1.5 hours in total. In each condition there were 10 critical signals per 10-min period on watch. In the high-event-rate condition stimuli appeared at an event rate of 59.2/minute. In the low event rate condition events occurred at a rate of 12.5 per minute. The task required observers to monitor the state of a synthetic system consisting of two inputs and an output, as shown in Figure 1. The stimuli consisted of three black dots and three rectangles. Each black dot was positioned above a rectangle. The input and output values were represented as the vertical distance of each .6 cm diameter black dot from its respective rectangle and observers monitored changes in these relative distances. Neutral events, requiring no overt response from the participant, were defined as cases in which each dot was a standard distance (.7 cm) from its rectangle. Critical signals for detection were cases in which any one of the dots was .5 cm closer than the standard distance to the rectangle. There were no cases in which more than one dot was closer than the standard distance. These stimuli have been used in a previous experiment and were associated with a demonstrated performance decrement (Szalma, 2002).

Following instructions regarding the nature of the task and the stimuli, participants in the choice conditions (matched and unmatched) were informed that they could experience either an easy or hard version of the task. The individuals in the no-choice group were not provided this opportunity. A 5-minute practice session was provided in which the participants experienced the event rate condition to which they were assigned for the vigil.

Perceived Workload and Stress

Perceived mental workload was measured by a computerized version of the NASA Task Load Index (TLX; Hart & Staveland, 1988). The standard version of the TLX was used, in which participants first provided ratings on each of the six subscales and then engaged in a paired-comparison procedure to determine the relative contributions of the subscales to the overall workload score. Perceived stress was measured using the Dundee Stress State Questionnaire (DSSQ), a well-established multidimensional measure of stress that assesses transient states associated with mood, arousal, and fatigue (Matthews et al., 1999). Specifically, the DSSQ provides three scale scores regarding the affective (*Distress*), motivational/cognitive (*Task Engagement*), and cognitive (*Worry*) components of stress state (Matthews et al., 2002). The NASA-TLX was administered after the vigilance session. The DSSQ was administered in two parts: a pre-vigil questionnaire completed prior to the initiation of the vigil, and a post-vigil questionnaire completed after the vigil. The order in which these were administered was counterbalanced across participants.

EXPERIMENTAL RESULTS

Data for all dependent measures were analyzed via a 2 (event rate) by 3 (choice) by 4 (period) mixed ANOVA with repeated measures on the last factor. For all tests involving repeated measures, Box's epsilon was used to adjust the degrees of freedom for violations of sphericity (Maxwell & Delaney, 2004). For the analyses of performance, the hit and false alarm rates were converted via arcsine transform for the inferential tests. The descriptive statistics are reported as untransformed proportions.

Correct Detections

The ANOVA revealed statistically significant main effects for choice, $F(2,66)=57.63$, $p<.001$, event rate, $F(1,66)=13.35$, $p<.01$, and period on watch, $F(2,115)=7.00$, $p<.01$. The three-way interaction among condition, event rate, and period on watch was also statistically significant, $F(3,115)=2.935$, $p<.05$. To further explore this interaction separate choice condition by period ANOVAs were computed within each event rate condition. At the higher event rate, statistically significant effects were observed for choice, $F(2,33)=20.93$, $p<.001$, and period on watch, $F(2,66)=4.63$, $p<.05$. The interaction between these factors failed to reach statistical significance ($p=.57$). These results are shown in Figure 2. Tukey HSD tests indicated that observers who experienced the condition matching their preference achieved higher detection rates (M=0.77) than those who whose group assignment was mismatched (M=0.23), or the no-choice control condition (M=0.40). Although the latter two conditions did not differ significantly from one another ($p=.13$), it is noteworthy that providing choice but assigning the individual to a mismatched condition impairs performance more than providing no choice at all (pairwise effect size, d=.29). With respect to the effect of period on watch, correct detections declined over time (M's=0 .50, 0.52, 0.45, and 0.41 for periods 1-4, respectively).

At the low event rate, a statistically significant effect was observed for the choice condition, $F(2,33)=59.44$, $p<.001$, and a marginal effect for period (after correction for sphericity), $F(2,51)=3.16$, $p=.06$. These effects are shown in Figure 3. Tukey HSD tests for the choice conditions indicated that observers in the choice-matched condition (M=.85) achieved significantly higher detection scores than their cohorts in

either of the other two groups, and that those in the no-choice condition (M=.65) achieved higher detection scores than those in the choice-mismatched group (M=.34). The interaction between choice condition and period on watch was also statistically significant, $F(2,51)=2.76$, $p=.05$. Tests for the simple effects of period within each choice condition indicated that detections declined with time on watch in the choice-mismatched condition, $F(1,14)=4.52$, $p<.05$, but not in the choice-matched or the no-choice conditions ($p>.50$ in each case). In sum, providing participants with 'choice' and then mismatching them with respect to their preference induces a steeper performance decrement as well as lower overall performance relative to a condition in which preference and group assignment were matched. Note that this was true regardless of whether the individual experienced a high or low event rate (i.e., whether they preferred an 'easy' or 'difficult' task).

False Alarms

The ANVOA indicated a significant effect for event rate, $F(1,66)=19.73$, $p<.001$, and period on watch, $F(3, 175)=11.73$, $p<.01$. All other sources of variance did not reach conventional levels of statistical significance ($p>.32$ in each case). As expected, there were more false alarms committed at the high event rate (M=.11) than at the low event rate (M=.02), and for both conditions false alarms declined with time on watch (M's=.08, .07, .06, and .06 for periods 1-4, respectively).

Perceived Workload

A 2 (event rate) by 3 (choice condition) ANOVA of global workload revealed a significant effect for choice, $F(2,66)=6.266$, $p<.05$. The effects for event rate and the interaction between choice and event rate were not statistically significant. Tukey HSD tests indicated that observers in the choice-mismatched condition reported significantly lower workload (M=52.025) than those in either the choice-matched (M=67.053) or no-choice (M=69.150) groups. The perceived workload of the latter two groups did not differ significantly from one another.

An ANOVA of the NASA TLX weighted subscales revealed a significant effect for scales, $F(3, 218)=12.74$, $p<.001$ and for choice, $F(2,66)=6.35$, $p<.01$. All other sources of variance were not statistically significant. However, across all conditions the profile of workload was consistent with prior research (Warm, Dember, & Hancock, 1996). Thus, Mental Demand (M=219.65), Temporal Demand, (M=223.47), and Frustration (M=210.90) contributed most to the workload of the task.

Stress

The pre-post change scores for the three secondary factors of the DSSQ were subjected to a 2 (event rate) by 3 (choice condition) ANOVA. Neither *Distress* nor *Worry* showed statistically significant pre-post differences across experimental conditions. For *Task Engagement* a significant effect was observed for choice, $F(2,66)=3.02$, $p=.05$. Although pre-post *Task Engagement* declined for participants in each of the three choice conditions, Tukey HSD tests indicated that the decline in *Task Engagement* in the choice-mismatched condition (M= -.80) was smaller than that associated with the no-choice group (M= -1.46). The decline in task engagement for observers in the choice-matched condition (M= -1.27) did not differ significantly from the scores of those in the other two groups.

DISCUSSION

Consistent with previous experiments, a higher event rate was associated with poorer performance (Warm & Jerison, 1984), although the concomitant effect on workload was not observed here as it had been in previous procedures (Warm et al., 1996). In addition, providing a choice attenuated the vigilance decrement, but this did depend on the particular event rate and whether the group assignment either matched or did not match the person's preference. At the higher event rate individuals who had a choice and experienced that condition still exhibited a decrement in detections, but their overall detection scores exceeded those in either of the other two groups. For the low event rate condition, allowing choice and matching individuals to that choice improved overall detection performance and attenuated the vigilance decrement. In contrast, providing the illusion of choice induced performance that was worse than those who had no choice at all.

A striking result was that, at the low event rate, providing a choice and then assigning the individual to a condition incongruent with that choice resulted in performance that was worse that that of individuals in the control condition who did not have the opportunity of choice at all. A similar but non-significant trend was observed for the high event rate, although the effect size was in the small to medium range (d=.29). Thus, the choice-mismatched condition impaired performance *regardless* of event rate. That is, regardless of whether the individual chose 'easy' or 'hard' or whether that was actually a low or high event rate, those in the choice mismatch condition performed more poorly. Hence, providing choice must be a real and not an illusory effect, and at low event rates mismatching may actually make things worse than providing no choice at all! It is particularly interesting that this effect was stronger at low event rates, since a mismatch in that condition means that the observers preferred a more difficult task. We assume that such participants wanted the challenge, and when they did not receive it they performed more poorly than those who did. These results partially confirm our hypotheses, in that the facilitative effect of choice on performance was greater when the group assignment was matched to the individual's preference. However, the expectation that these effects would be more pronounced in the high event rate condition was not confirmed. Indeed, it was at the low event rate that the larger differences were observed (see Figures 2 and 3).

Recall that the lower detection rates associated with the choice-mismatched condition were accompanied by lower

perceived workload relative to the no-choice control condition. This represents a performance-workload dissociation, in which the linkage between performance and workload is in the direction opposite to that expected (Hancock, 1996; Yeh & Wickens, 1988). Hancock (1996) observed that these patterns of dissociation can be diagnostic with respect to how observers deal with task demands. The pattern observed in the current study is consistent with a disengagement from the task among observers assigned to the choice-mismatched group. That is, participants who were mismatched 'gave up' and diverted effort away from task performance (cf., Hockey, 1997). Again, it is noteworthy that this pattern occurred across event rate conditions, indicating that participants who were mismatched gave up *regardless of the actual difficulty of the task*. However, the disengagement hypothesis is not consistent with the findings in regard to the DSSQ, that the decline in *Task Engagement* was *smaller* for the choice-mismatched condition relative to the other two groups. One possibility is that the lower perceived workload may have derived from greater *Task Engagement* on the part of these participants, but that this commitment of effort was not sufficient to overcome the performance effects of mismatching their task experience with their preferences. The reason for the failure of higher levels of effort to compensate for mismatching preferences is a matter for future research. However, it may be that mismatching preferences diverts cognitive resources away from the task by directing attention to the violation of expectancy regarding 'task difficulty,' thereby forcing the task into a data-limited region in which further allocation of effort does not improve performance (Norman & Bobrow, 1975).

In sum, the results of the current study underscore the importance of motivational effects in vigilance in general and, in particular, the assertion by Dember and his colleagues (1992) that choice is a crucial element in vigilance.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. P.A. Hancock*, Principal Investigator.* The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant. The authors also thank H.C.N. Ganey for assistance in data collection.

REFERENCES

Davies, D.R., & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Dember, W.N., Galinsky, T.L., & Warm, J.S. (1992). The role of choice in vigilance performance. *Bulletin of the Psychonomic Society*, *30*, 201-204.

Gunn, D.V., Warm, J.S., Nelson, W.T., Bolia, R.S., Schumsky, D.A., & Corcoran, K.J. (2005). Target acquisition with UAVs: Displays and advanced cueing interfaces. *Human Factors*, *47*, 488-497.

Hancock, P.A. (1996). Effects of control order, augmented feedback, input device, and practice on tracking performance and perceived workload. *Ergonomics*, *39*, 1146-1162.

Hancock, P.A. (1998). The price of freedom. *Proceedings of the Human Factors and Ergonomics Society*, *42*, 1577-1578.

Hancock, P.A., & Szalma, J.L. (2003). Vigilance and the price of freedom. *HSIAC Gateway*, *13*, 5, 20.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P.A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive –energetical framework. *Biological Psychology*, *45*, 73-93.

Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., & Warm, J.S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, *2*, 315-340.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state 'big three'? In: I. Mervielde, I.J. Deary, F. DeFruyt, and F. Ostendorf (Eds.), *Personality psychology in Europe (vol 7)* (pp. 335-350). Tilburg: Tilburg University Press.

Maxwell, S.E., & Delaney, H.D. (2004). *Designing experiments and analyzing data: A model comparison perspective, 2nd ed.* Mahwah, NJ: Erlbaum.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processing. *Journal of Cognitive Psychology*, *7*, 44-60.

Scerbo, M.W. (1998). Fifty years of vigilance research: Where are we now? *Proceedings of the Human Factors and Ergonomics Society*, *42*, 1575-1576.

See, J.E., Howe, S.R., Warm, J.S., & Dember, W.N. (1995). A meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin*, *117*, 230-249.

Szalma, J.L. (2002). Workload and stress of configural displays in vigilance tasks. *Proceedings of the Human Factors and Ergonomics Society*, *46*, 1536-1540.

Szalma, J.L., Hancock, P.A., Dember, W.N., & Warm, J.S. (2006). Training for vigilance: The effect of KR format and dispositional optimism and pessimism on performance and stress. *British Journal of Psychology*, *97*, 115-135.

Szalma, J.L., Warm, J.S., Matthews, G., Dember, W.N., Weiler, E.M., Meier, A., & Eggemeier, F.T. (2004) Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Human Factors*, *46*, 219-233.

Warm, J.S., & Dember, W.N. (1998). Tests of vigilance taxonomy. In: R.R. Hoffman, M.F. Sherrick, and J.S. Warm (Eds.). *Viewing psychology as a whole: The integrative science of William N. Dember*. (pp. 87-112). Washington, DC: American Psychological Association.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In: R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183-200). Hillsdale, NJ: Erlbaum.

Warm, J.S., & Jerison, H.J. (1984). The psychophysics of vigilance. In: J.S. Warm (Ed.), *Sustained attention in human performance*. (pp. 15-59). Chichester, United Kingdom: Wiley.

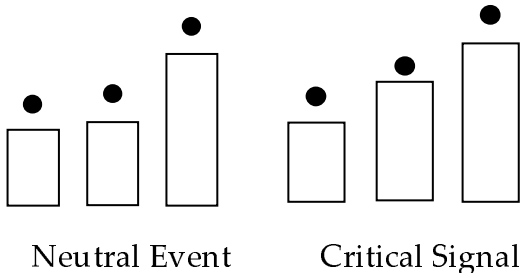


Figure 1. Example stimuli employed for the present study.

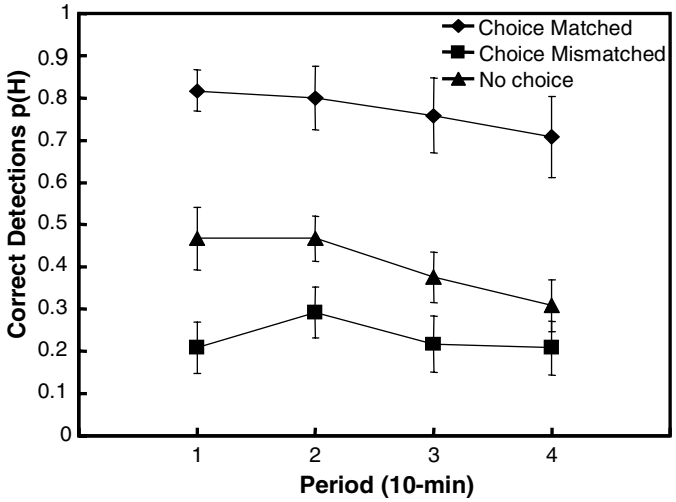


Figure 2. Proportion of correct detections as a function of period on watch at high event rate. *Note*. Error bars are standard errors.

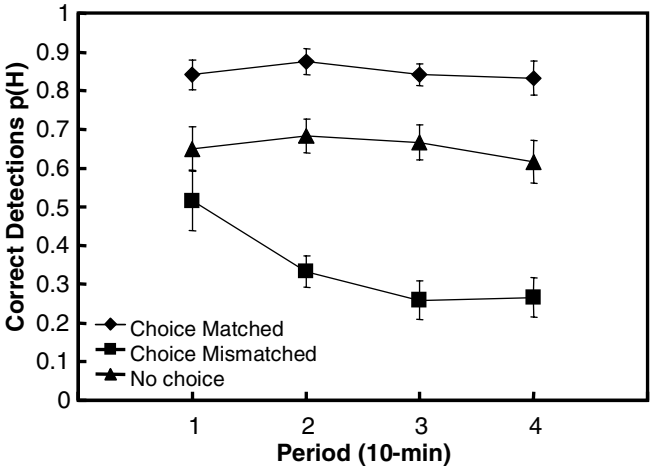


Figure 3. Proportion of correct detections as a function of period on watch at low event rate. *Note*. Error bars are standard errors.

Low and high levels of attentional control in adaptive levels of automation

J.E. Thropp, T. Oron-Gilad, J.L. Szalma, and P.A. Hancock
MURI Laboratory
University of Central Florida

Abstract

Adaptive automation that functions according to user individuation has many potential benefits, and is an area that warrants investigation. A strong understanding of human individual differences can therefore serve as an important foundation for developing an adaptive, user-centered form of automation that best accommodates human-machine interaction. The role of individual differences in attentional control upon preference of automation level was assessed. Participants with low and high attentional control were allowed to choose among four LOAs to complete visual and auditory target detection tasks. A trend for attentional control level to be inversely proportional to LOA preference was found.

Introduction

Individual differences play a dominant role in determining how well a job can be performed, accounting for more variability in performance than differences in systems designs and/or training procedures (Egan, 1988). Yet, many computerized systems have been developed without individual differences in capability. A strong understanding of human individual differences can therefore serve as an important foundation for developing an adaptive, user-centered form of automation that best accommodates human-machine interaction.

Many individual difference variables should be investigated (Thropp, Oron-Gilad, Szalma, & Hancock, manuscript submitted); variability in attentional control has been chosen based upon previous findings that strongly extend to the adaptive automation domain.

Attentional control is important for the ability to orient one's attention to the appropriate aspect of an environment (Derryberry & Reed, 2002). Attentional focusing and shifting measures have been combined into a new scale of Attentional Control (Derryberry & Reed, 2002), which measures a general ability to orient one's attention to the appropriate aspect of an environment as necessary. The Attentional Control Scale has specific subfactors for focusing attention, shifting attention between tasks, and controlling thoughts. Attentional control (AC) scores can range from 20 (low) to 80 (high).

The user with high attentional allocation can keep track of these different tasks and thereby complete multiple tasks simultaneously. High attentional control may therefore be conducive to more manual control of system functions, although this highly adaptive characteristic would likely enable the individual to operate successfully at any LOA. Operators with lower attentional control may be facilitated by higher LOAs, as the

operator may have deployed these resources to other environmental stimuli instead of task-relevant stimuli.

Method

Experimental Participants

Thirty-six participants (28 females, eight males) were recruited from a subject pool of 303 psychology undergraduates, all of whom completed the Attentional Control Scale (Derryberry & Reed, 2002). They ranged in age from 18 to 34 (M = 22.25, SD = 3.524).

To separate 'low' and 'high' AC scores and maximize variance between the groups, a quartile split was performed upon the AC scores from the original subject pool using SPSS. This procedure indicated that the lower-bound of the upper quartile was an AC score of 56. Conversely, the upper-bound of the lower quartile was an AC score of 44. Thus, the high AC category was determined by a AC scores ranging from 56 to 80, while the low AC scores ranged from 20 to 44. The AC scores for the 'Low AC' group ranged from 36 to 44 (M = 39.94, SD = 2.461). The AC scores for the 'High AC' group ranged from 56 to 73 (M = 60.22, SD = 4.209).

The Automation Program

Each participant undertook the role of the supervisor of a machine decision-aid which had 90% accuracy in correctly identifying targets and nontargets in a series of perceptual tasks. They were given the choice of four different LOAs in which to complete the perceptual tasks, based upon both their preferences for operator-machine interaction and their needs in order to assure accurate target identification. Participants identified each stimulus type after it was presented by using a mouse to click on the buttons on the display labeled 'Target' and 'Non'.

In the visual-only target detection task, participants differentiated between black vertical lines (32mm in height) presented at either a short (125 ms) or long (250 ms) period of time upon a white background. Targets were lines appearing for the shorter period of time, while nons were lines appearing for the longer period of time. Target and non lines as well as target and non tones were presented in random orders. During the visual-only task, the auditory task was fully executed by the computer, and the participant was instructed to ignore the auditory task.

In the auditory-only target detection task, participants distinguished between sounds (400 Hz C4 triangle waveform; see Droit-Volet, 2004) that were either short (200 ms) or long (250 ms) in presentation time. During the auditory-only task, the visual detection task was fully executed by the computer, and the participant was asked to ignore the visual task.

In the audiovisual combined task, both the audio and visual target detection tasks were presented simultaneously, and the participant and machine were responsible for performing none, one, or both tasks according to the participant's chosen LOA and method of operating the LOA. As in the visual-only and auditory-only tasks, both lines and tones were presented in random orders, but did not overlap temporally or spatially such that only one stimulus was presented at a time.

A temporal discrimination task was chosen because stimulus presentation time is a characteristic that can be manipulated in both auditory and visual displays. These visual and auditory short and long temporal values have been psychophysically equated for discriminability (see Szalma et al., 2004).

Participants were able to change the level of automation (LOA) according to their preference throughout the course of the task. There were four LOAs available to the participants: '1 Full Operator', '2 Mostly Operator', '3 Equal Sharing, and '4 Mostly Machine', as described in the methodology. In LOA '1 Full Operator' work was conducted solely by the operator without the machine aid. LOAs 2, 3, and 4 provided the machine decision aid. In LOA 2, the operator responded first and then received a machine suggestion. In LOA 3, the operator responded first to all stimuli in the left field and then received a machine suggestion, while in the right field, the machine responded first and the operator could veto the machine decision. In LOA 4, the machine responded first and then the operator had the option to input her own response if it differed from the machine response. Participant's preferences for the different LOAs were assessed by examining the percentage of time that participants chose to be in each one of the LOAs, relative to the entire duration of the task.

Four separate tasks were analyzed: visual-only and auditory-only, both of which were the single-task conditions; and the visual portion of the audiovisual-combined task (referred to as 'visual-combined') and the auditory portion of the audiovisual combined task (referred to as "auditory-combined"), both of which were the dual-task conditions. Thus, two load conditions were assessed. Single-task conditions encompass the visual-only and visual-combined tasks, and dual-task conditions encompass the visual-combined and auditory-combined tasks (i.e., the entire audiovisual combined task).

Two modalities were differentiated: visual and auditory. Furthermore, we analyzed the responses of the operator in relation to the decision aid (without and with machine aid) hence there were two levels of responses: operator-alone and machine-aided. Operator-alone involved the participants' responses independent of the machine aid, while machine-aided involved the participants' final responses after using the machine decision-aid.

Results

Preferences for LOAs

A 2 (Modality) x 2 (Load) x 4 (LOA) within-subjects repeated measures ANOVA was conducted. AC was a between-subjects variable.

There was a significant main effect for LOA, $F(3, 105) = 6.841$, $p < .0001$. Post hoc tests conducted with Tukey's HSD ($q = 3.825$, $t = 2.696$) revealed that that significantly more time was spent at levels '1 Fully Operator' than '3 Equal Sharing', $t(35) = 3.257$, $p < .005$. Significantly more time was spent at '2 Mostly Operator' than '3 Equal Sharing', $t(35) = 5.263$, $p < .0001$. Significantly more time was also spent at '4 Mostly Operator' than '3 Equal Sharing', $t(35) = -4.862$, $p < .0001$.

The mean proportion of time spent at LOA '1 Full Operator' was .218 (*SD* = .346), .397 at '2 Mostly Operator' (*SD* = .420), .024 (*SD* = .070) at '3 Equal Sharing', and .361 (*SD* = .414) at '4 Mostly Machine'. See Figure 1 for a depiction of the time proportions.

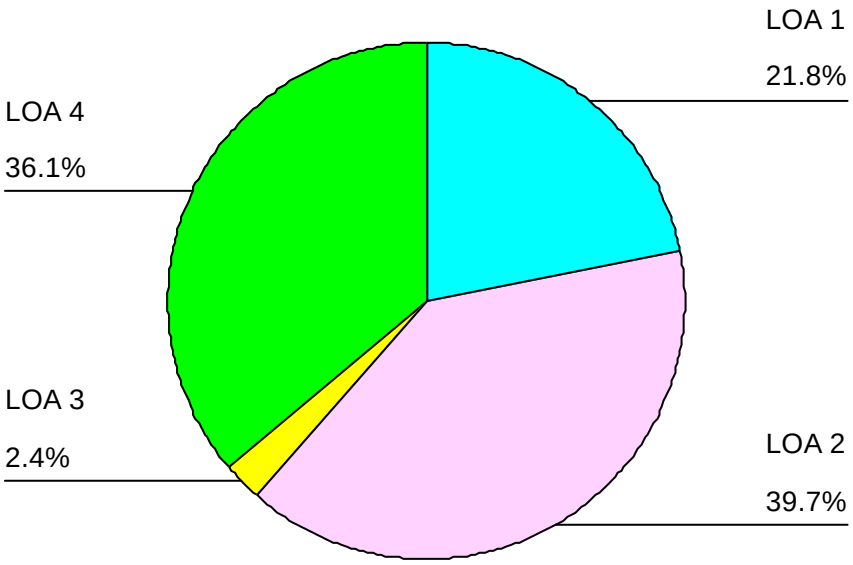


Figure 1. Proportion of time spent at each LOA collapsed across tasks.

Although there was no significant interaction between LOA and AC, $F(2, 68) = .587$, $p > .05$, there was a trend for those with low AC to use higher LOAs than those with high AC across all four tasks (see Figure 2). A priori test of simple effects indicated that those with low AC spent significantly more time at LOA 4 than LOA 1 (M = .482, SD = .467, and M = .096, SD = .234, respectively), $t(17) = 2.851$, $p < .05$. There was also a trend for those with low AC individuals to spend more time at LOA 4 than those with high AC (M = .441, SD = .377, and M = .291, SD = .447, respectively). Conversely, those with high AC spent more time at LOA 1 than those with low AC (M = .304, SD = .419, and M = .138, SD = .237, respectively).

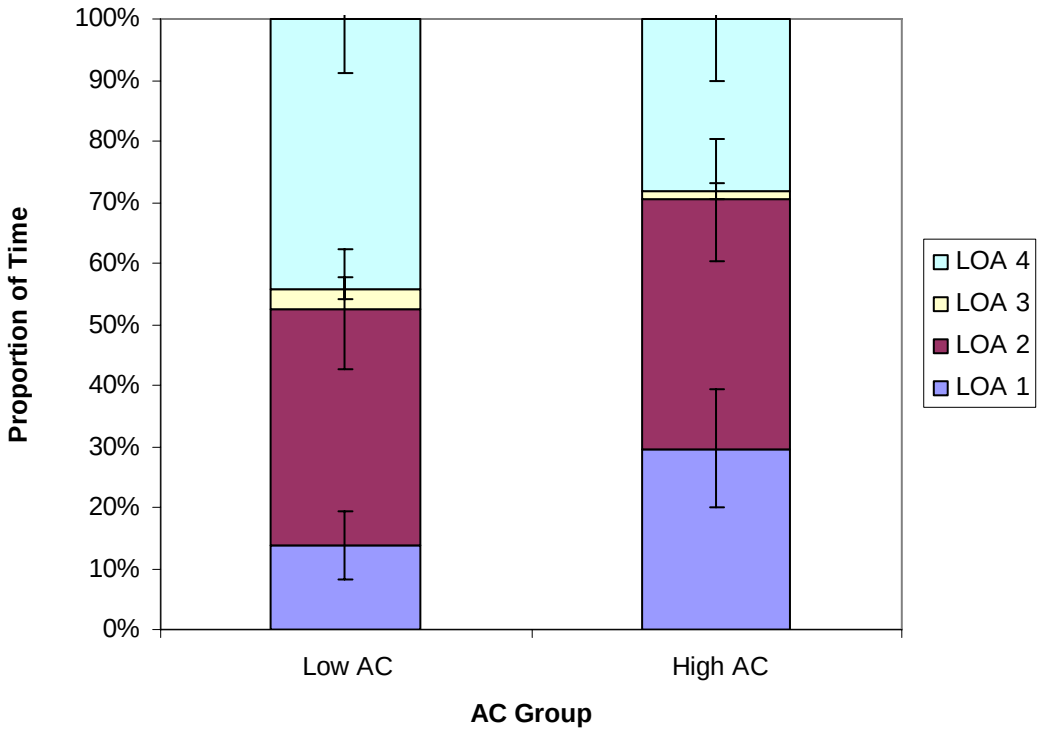


Figure 2. Overall proportion of time spent at each LOA collapsed across tasks for low and high AC groups.

Signal Detection Performance - d' (Sensitivity)

d' is a measure of sensitivity, or the participant's ability to correctly identify the stimuli as a targets or nons, and is one of the most commonly used measured of sensitivity in detection theory (Green & Swets, 1966). High sensitivity is indicated by a d' which is greater than 0. Low sensitivity is characterized by a hit rate and false alarm rate that approximate equivalency and is indicated by a d' that approximates 0.

A 2 (Modality) x 2 (Load) x 2 (without-with Machine Aid) within-subjects repeated measured ANOVA was conducted. AC was the between-subjects variable.

There was a main effect for modality, such that the d' in the auditory tasks (M = 2.163, SD = .572) was significantly higher than the d' for the visual tasks (M = 1.143, SD = .552), $F(1, 34) = 52.343$, $p < .0001$. Post-hoc comparisons between tasks were then made. Tukey's HSD (q = 4.563, t = 3.227) revealed that there was a significant interaction in the operator-alone d' between the visual-only and auditory-only tasks $t(34) = 6.099$, $p < .05$, such that the operator's d' of the auditory-only task was higher than the operator's d' of the visual-only task (M = 2.036, SD = .991, and M = .902, SD = .779). In the dual-task condition, the operator-alone d' of the auditory-combined task was significantly higher than that of the visual-combined task (M = 2.193, SD = .898, and M = 1.169, SD = .735, respectively), $t(34) = 5.554$, $p < .05$.

There was a main effect for use of the machine's decision-aid, as it significantly improved d' across tasks and modality, $F(1, 34) = 4.651$, $p < .05$. Across the visual

modality, there was a significant effect for use of the machine decision aid, $t(35) = 2.141$, $p < .05$, such that the operator's d' was significantly when the machine was used ($M = 1.251$, $SD = .589$) than the operator's d' without the machine aid ($M = 1.036$, $SD = .659$). Across the auditory modality, however, there was no significant improvement for using the machine aid ($p > .05$).

For participants with high AC, the overall d' in the auditory-combined task was significantly higher than that of the auditory-alone task ($M = 2.311$, $SD = .764$, and $M = 1.931$, $SD = .449$, respectively), $t(17) = 2.183$, $p < .05$. The overall d' in the auditory-only task was also significantly higher than the d' in the visual-only task ($M = 1.931$, $SD = .764$, and $M = .951$, $SD = .614$, respectively), $t(17) = 4.119$, $p = .001$. The overall d' in the auditory-combined task was also significantly higher than the visual-combined task d' ($M = 2.311$, $SD = .449$, and $M = 1.173$, $SD = .618$, respectively), $t(17) = 5.897$, $p < .0001$.

For participants with low AC, The overall d' in the auditory-only task was significantly higher than that of the overall visual-only task ($M = 2.357$, $SD = .766$, and $M = 1.077$, $SD = .453$, respectively), $t(17) = 6.959$, $p < .0001$. The overall d' in the auditory-combined task was also significantly higher than that of the visual-combined task ($M = 2.060$, $SD = 1.059$, and $M = 1.373$, $SD = .756$, respectively), $t(17) = 2.406$, $p < .05$. This did not exceed Tukey's critical t of 3.227, however, the Cohen's d was .747.

Response Bias - β (Beta)

β refers to the operator's response bias in identifying stimuli as targets or nons. Macmillian and Creelman (2005), refer to it as a particular favoritism towards a response type. The value of β ranges from 0 to greater than 1. Values in which $\beta > 1$ indicate that the individual tends to classify more stimuli as nons (the miss rate exceeds the false alarm rate), resulting in more misses than hits. Conversely, values in which $\beta < 1$ indicate that the individual tends to classify more stimuli as targets (the false alarm rate exceeds the miss rate), thereby resulting in more false alarms than hits.

A 2 (Modality) x 2 (Load) x 2 (Machine Aid) within-subjects repeated measures ANOVA was conducted. AC was a between-subjects variable.

There was a significant main effect for AC, $F(1, 34) = 8.185$, $p < .01$, such that participants with high AC had a significantly higher mean β than those with low AC ($M = 1.286$, $SD = .491$, and $M = .920$, $SD = .229$, respectively). More specifically, participants with high AC had a significantly higher mean operator-alone β than those with low AC, $t(34) = 3.361$, $p < .005$, ($M = 1.4364$, $SD = .620$ and $M = .908$, $SD = .244$, respectively). There was no significant difference in machine-aided β between those with low and high AC ($p > .05$).

There as a significant interaction between AC and modality, $F(1, 34) = 6.408$, $p < .05$. Tukey's HSD ($q = 3.820$, $t = 2.701$) revealed that in the auditory modality, those with high AC had a significantly higher mean β than those with low AC ($M = 1.570$, $SD = .749$, and $M = .940$, SD = .386, respectively), $t(34) = 3.169$, $p < .005$. However, there was no such difference within the visual modality $t(34) = .822$, $p > .05$. See Figure 3 for a depiction of the mean β in each modality grouped by AC.

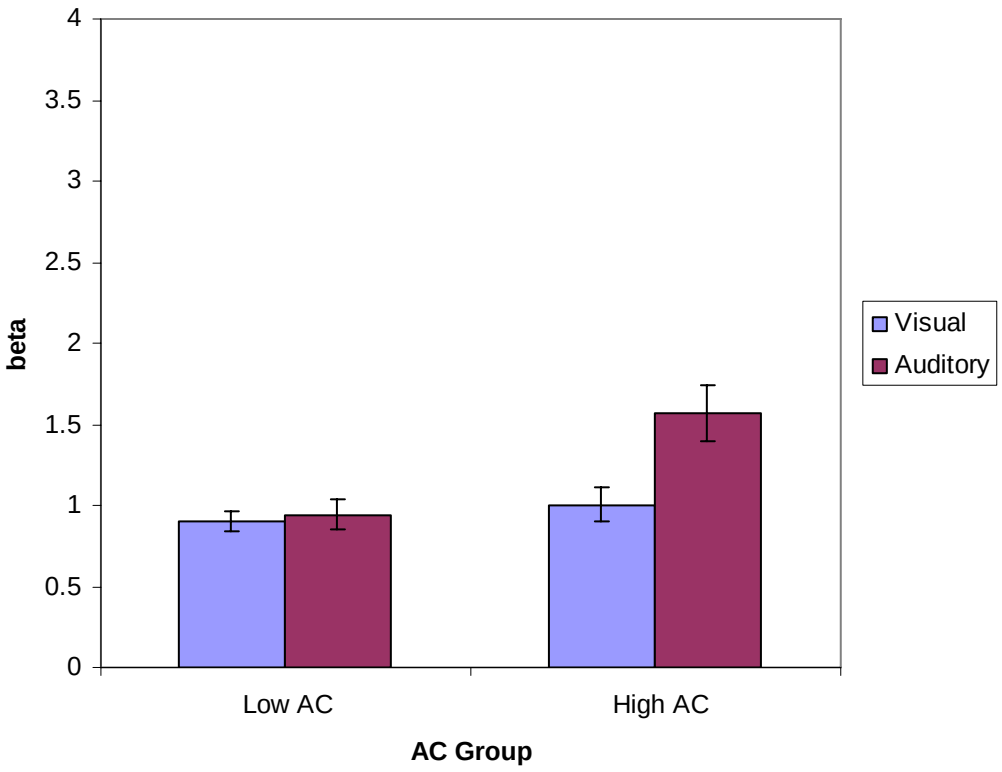


Figure 3. Mean β by AC group and modality.

Subjective Workload Ratings

Workload measurements were taken three times over the course of the experiment: after the visual only condition, after the auditory condition and after the combined condition. A 3-way (Task) within-subjects repeated measures ANOVA was used. AC was a between subjects variable.

A priori tests of simple effects found that the high AC group reported significantly higher overall workload in the visual-only task than the low AC group (M = 51.37, SD = 11.948, and M = 41.95, SD = 12.909, respectively), $t(34) = 2.271$, $p < .05$. Although not significant, there was a trend for the high AC group to report higher overall workload in the auditory-only task (M = 46.97, SD = 7.710, and M = 44.07, SD = 11.597, respectively) and in the combined task (M = 53.94, SD = 48.19, SD = 13.858, and M = 48.19, SD = 13.180, respectively). See Figure 4 for a graph representing the overall workload reports in each task by AC group.

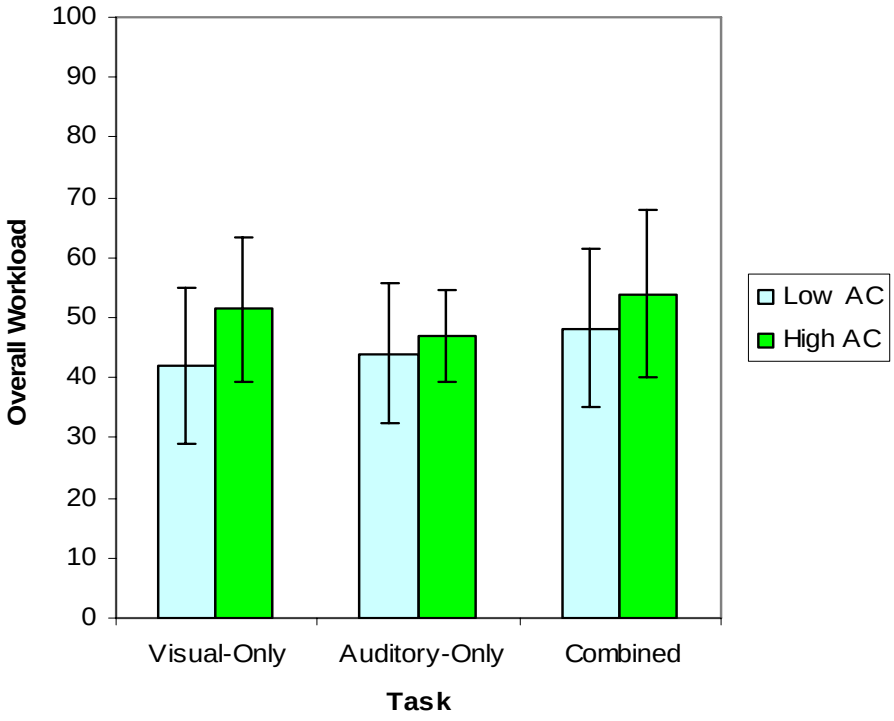


Figure 4. Mean overall workload rating in each task for low and high AC groups.

Discussion

Statistically, LOAs 1, 2, and 4 were equally preferred across tasks and levels of AC. Thus, participants from both AC groups found them all to be equally usable and useful. Participants equally preferred various amounts of assistance from the machine, ranging from no assistance (i.e., LOA 1), to supervising the machine as it made all of the stimuli discriminations and intervening only when a machine decision error was suspected (i.e., LOA 4). LOA 2 was also frequently chosen; participants may have preferred completing the tasks at this level as they received a type of performance feedback, even if 90% accurate.

Participants of both the low and high AC groups spent minimal time at LOA 3, Equal Sharing, when given the choice of four different LOAs. LOA 3 required even less operator responsibility than LOAs 1 and 2, both of which were more preferred than LOA 3. It was also less popular than LOA 4, in which minimal operator responsibility was required as a supervisory role was assumed. This suggests that LOA 3 may have been least preferred for reasons other than workload and responsibility. One possible reason may be that operators did not wish to filter stimuli, such that they were required to attend to certain stimuli (i.e., left fields) while ignore others of equal perceptual intensity and modality type (i.e., right fields).

No significant main AC effects were found in terms of LOA preference, however, there were trends for those with low AC to spend more time at LOA 4 than those with high AC, collapsed across tasks. Also, those with high AC spent more time at LOA 1 than did those with low AC. Thus, lower AC levels were conducive to the preference for more automation and higher AC levels were conducive to inclinations towards less automation. This supports the prediction that AC is inversely proportional to automation use. However, the preference for LOA 2 was roughly equivalent for both low and high AC groups; indicating that both types of individuals liked using LOA 2, a moderately-low amount automation.

Although not hypothesized, participants were generally more sensitive to auditory stimuli than visual stimuli, as evidenced by the higher d' for the auditory tasks. Despite prior psychophysical equation of these two discrimination types by Szalma et al. (2004), the auditory discriminations were easier than the visual discriminations. This modality effect in d' also replicated the results by Szalma et al. 2004). Additionally, this modality d' difference persisted in individuals with both low and high AC.

Significantly higher reports among the high AC participants occurred in the overall workload rating in the visual-only task, and effort in the visual-only task. The reasons for this may include the tendency for those with high AC to spend more time in LOA 1 than those with low AC. Conversely, the low AC participants spent more time in LOA 4 overall. Another possible explanation for this is that the high AC participants may have been attending to the nature of the task more so than the low AC participants. Thus, the high AC group may have been more engaged in the tasks, and therefore more prone to noticing their own feelings of workload.

While three of the available LOAs were equally preferred, and task performance was variable and neither subjected to floor nor ceiling effects, operators may not have necessarily accurately judged their need for various LOAs. In this experiment, the aim was determining preference rather than need for automation. However, a discrepancy between need and preference has been noted by Morrison and Rouse (1986), as the operator's ability to assess the need for automation may not always be realized.

References

Derryberry, D. & Reed, M.A. (2002). Anxiety-related attentional biases and their regulation by attentional control, *Journal of Abnormal Psychology*, *111*(2), 225-236.

Droit-Volet, S., Tourret, S. & Wearden, J. (2004). Perception of the duration of auditory and visual stimuli in children and adults. *The Quarterly Journal of Experimental Psychology,57A*(5), 797-818.

Egan, D. E. (1988). Individual differences in human-computer interaction. In Handbook of Human-Computer Interaction (M. Helander, Ed.) Elsevier Science Publishers B.V. (North-Holland).

Morrison, N.M. & Rouse, W.B. (1986). Adaptive Aiding for Human-Computer Control: Experimental Studies of Dynamic Task Allocation. (Tech. Rep. AAMRL-TR-86-005). Naval Air Warfare Center, Warminster, PA.

Szalma, J.L., Warm, J.S., Matthews, G., Dember, W.N., Weiler, E.M., & Meier, A. (2004). Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Human Factors, 46*(2), 219-233.

Thropp, J.E., Oron-Gilad, T., Szalma, J.L. & Hancock, P.A. Incorporating individual differences into adaptive automation. Manuscript submitted.

DEVELOPMENT OF AN OPERATIONAL SETTING SPECIFIC FIELD COGNITIVE ASSESSMENT PROCEDURE

Wayne C. Harris[1], Dennis L. Reeves[2], Timothy F. Elsmore[3] and Peter A. Hancock[1]

The assessment of dynamic cognitive abilities during actual field operations by professional personnel represents a significant challenge to the behavioral scientist. Here, we present a short-duration, portable cognitive assessment instrument which seeks to address such issues. Any such instrument with poor face validity is unlikely to receive the required attention and cooperation of operational personnel. Also, if supervisors see little relationship between the assessment procedure and the actual operational tasks they are unlikely to use outcome results anyway. This work summarizes our customization of an established cognitive assessment battery, which is composed of operationally relevant tasks for field assessment. Our present results suggest equivalent user acceptance and comparable assessment capacity as the longer duration, currently established battery.

[1] University of Central Florida, [2] Clinvest, Inc., [3] Activity Research Services

INTRODUCTION

Technology now performs tasks that in the past required human operators to accomplish. Technology further provides decision support to human operators but system performance is still largely contingent upon the operator's ability to manage the system. The frequency of actual operator-machine interactions may have decreased, but cognitive demands are heightened as such periodic decisions assume ever greater importance, while additional supervisory responsibility also results in the need for additional cognitive effort (Sheridan, 1987). Sufficient cognitive ability for jobs is determined during the selection procedure and demonstrated and enhanced during training, but cognitive capacity is itself dynamic. External and internal conditions affect it and conditions that serve to decrease cognitive capacity are frequent in settings where cognitive deficits are often most costly. Individuals, and their supervisors, must continually evaluate cognitive capacity and adjust cognitive demands appropriately, replace impaired personnel from critical assignments and mitigate stress effects with interventions such as nutritional support and flexible work scheduling. However, any such intervention requires an accurate measure of current cognitive capacity, and especially in field conditions such an assessment represents a formidable challenge.

REVIEW

There have been several systematic attempts to generate these flexible and robust test capacities. For example, computer versions of classic tests of cognitive functioning are available in the Walter Reed Performance Assessment Battery (PAB) (Englund., et al, 1987). PAB task are classified by cognitive abilities which represent elemental skills that underlie real-world tasks (Thorne, Genser, Sing, & Hegge, 1985). The Automated Neuropsychological Assessment Metric (ANAM) (Reeves, et al., 1991) organized and presented PAB tasks in a user-friendly manner. The Automated Readiness Evaluation System (ARES) developed by Elsmore and Reeves (in press) is an extension of this battery and presents ANAM tasks on handheld computers. The ARES has proven to be a most useful method of assessing cognitive functioning in operational settings (Harris, & Hancock, 2003; Harris & Hancock, & Morgan, 2005), and component test profiles have been constructed for field settings that provide a profile of the cognitive

capacities deemed to be important in those settings (Harris, Hancock, & Harris, 2005; Harris, & Hancock, 2003; Harris, Hancock, & Morgan, 2005).

The ARES assesses cognitive capacities using a series of generic tasks and abstract symbols which are derived presently to represent recognizable elements of common military tasks, although in principle any such abstract representations could be developed. Observations across a series of tests may indicate that that cognitive capacity is declining, but changes can also indicate decreases in motivation. Performance during operational conditions is typically accompanied by high levels of motivation, and motivation should also be high during assessments to accurately estimate operational cognitive functioning. The more irrelevant an assessment activity appears the less effort an individual is likely to expend when their cognitive resources become depleted and as a consequence, the less accurately will assessments predict actual operational effectiveness. In the present version, care was taken to address this specific issue of motivation to perform versus actual capacity change.

DEVELOPMENT OF AN OPERATIONS TESTING RELEVANT COGNITIVE BATTERY

To disambiguate the potential confound between actual decrement and motivation level, a battery of established basic cognitive functions tests was assembled and then modified by changing the appearance of each task so as to be relevant to the current job performance in the field. We anticipated that this manipulation would sustain operational levels of motivation and that assessments would therefore more accurately reflect the person's current ability to perform cognitively demanding in the field.

In addition to the motivational value of the new task configurations it was postulated that assessments with little relationship to current assignments would be unlikely to be used by supervisory personnel. Cancellations and delays are usually costly, and an obvious relationship between an assessment and operational activities needs to be present for supervisors to justify operational changes.

Battery development began with the Automated Readiness Evaluation System (ARES), a cognitive assessment tool that includes tasks that are relevant to military performance and are sensitive to fluctuations in demand associated with infantry operations (Harris, Hancock & Morgan, 2005).

The ARES-I (ARES Infantry Version)

In ARES-I (Infantry Version), tasks were developed to assess Simple Reaction Time and Logical Relations and Spatial Processing. These basic cognitive performance capacities have been shown to be sensitive to infantry operational conditions (Harris, Hancock, & Morgan, 2005). Tests were modified to retain the psychological demands of the ANAM/ARES tasks, while presenting those demands in a task structure relevant to infantry personnel. Three specific ANAM tasks were modified.

1. Simple Reaction Time

The standard ARES includes a test that assesses Simple Response Time. A simple asterisk appears on the screen and the participant's task is to press a designated button as rapidly as possible. The ARES-I includes a task with similar cognitive demands, and adds two tasks in which responses require stimulus classification. Figure 1 presents the revised Simple Reaction Time display in the ARES-I. The participant presses a key when a stimulus (a white silhouette) appears. The timing of their appearances is identical to the ARES Simple Reaction Time task.

Decisions are frequently not limited to just detecting when an object appears. They require discriminating between objects and responding appropriately. Two ARES-I tasks were reconfigured that required such stimulus classification. In these tasks, white and green silhouettes appear in windows. In the first two-stimuli task, participants respond (by pressing the button labeled '5') only to white figures.. This task corresponds to a shoot/no shoot decision. In the second two-stimulus task, the participant presses the right key when a white

silhouette appears, and they are required to press the left key when the silhouette is green. Thus, participants classify all stimuli.



Figure 1. *ARES-I Reaction Time display.*

2. Logical Relations

The standard ARES abstract reasoning task is based on a task introduced by Baddeley (1968) in which two symbols appear on the screen (# &). A sentence above this display describes the relationship between the symbols such as "& is first." The participant's task is to indicate whether the statement is true (left button) or false (right button). In the ARES-I task two military vehicle map symbols move across the screen from left to right and below them a statement describing the relationship of the vehicles (Figure 2).

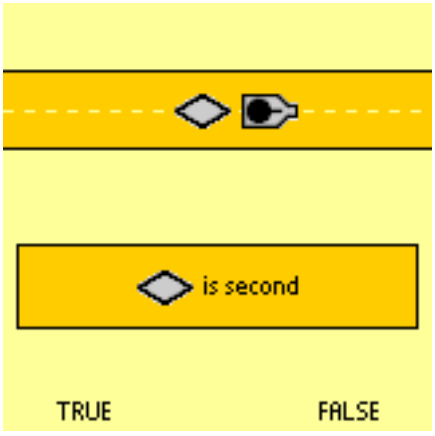


Figure 2. *ARES-I Logical Relations display.*

3. Spatial Processing

To evaluate spatial processing, the ARES figure rotation task uses 4x4 grids. Two grids are presented, one of which is rotated 90^0 and the participant's task is to indicate whether the figures are the same or different. In the ARES-I, a map with three dots representing military units appears (see Figure 3). The map disappears for 2 seconds after which it reappears rotated 90^0, 180^0 or 270^0. The participant's task is to indicate whether the unit locations are the same or different than their location on the first map. The map remains constant for all trials.

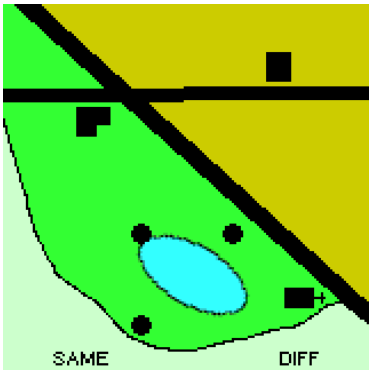


Figure 3. *Map rotation display.*

ARES-I EVALUATION

Initial ARES-I evaluation consisted of comparison of ARES-I results with previous ARES data and comments from users. Evaluations were conducted using police officers who completed the ARES-I four times during a day of weapons qualification, and Marine instructors who completed the ARES-I battery at the end of each workday for 30 days.

Evaluation of ARES-I included comparison of data with previous ARES field data and user comments. Response time changes during the first four assessments were consistent with previous ARES data. Accuracy was greater than 90% during the first trial and remained at that level during all four trials. Consistent with previous examinations of ANAM and ARES during initial testing

(Benedetto, Harris, & Goernert, 1995; Harris, Hancock, & Morgan, 2005), response time improved between trials one and four, with the rate of improvement decreasing with the number of trials and performance stabilizing on trial four.

Comparison of ARES-I response times with comparable data from Special Forces soldiers (Harris, Hancock, & Morgan, 2005) indicated that SRT mean response times were longer on the ARES-I. The longer response time may reflect smaller target ARES-I stimuli size, more complex background and moving stimuli location. Differences may also reflect the characteristics of the sample. Group means were Special Forces (220ms), Marine Instructors (255ms), and Police Officers (274ms).

The Convoy task response time and the relative time difference between Simple Response Time and Abstract Reasoning similarly suggest that the present interface changes did not significantly change the cognitive demands of the task. ARES-I response time differences in the three Reaction Time tasks increased with task complexity.

Consistent with previous findings that performance improves during trials when the tasks are being learned and increases with fatigue, within trial performance of police officers improved during the first trial. Within trial improvement ceased as the number of trials increased and within trial performance deteriorated during the fourth trial late in the day. The results are consistent with ANAM within trial data and most probably represents fatigue effects late in the day. (Harris, Hancock, & Harris, 2005)

The critical role of motivation was evident from the decrease in the number of Police Officers returning for assessments as the day progressed and the sporadic data from 3 of the 6 Marine volunteers. Police Officers indicated the tasks were similar to challenging games and appeared to be motivated. They were interested in their scores, particularly when testing occurred in group settings. Comments from the Marines indicated that assessment tasks were relevant to their work. Those who completed the battery indicated that it was challenging and even enjoyable at times but that the 15-minute completion time was a problem. Workdays often end between 2200 and 0200 hours thus adding 15 minutes to the workday required a major commitment. Negative comments were aimed primarily at the map task. The task required both spatial rotation and spatial memory, and its relatively high task difficulty was indicated by longer response times and error rates 3-4 times greater than other ARES-I tasks. Average Marine cognitive performance remained stable during 30 days of assessment. Day-to-day changes appear to reflect cognitive ability variability but no decrement indicative of loss of motivation was found.

DISCUSSION

Difficulty predicting real-world performance from cognitive assessments may be at least partially due to low motivation during assessment. To increase such motivation, an assessment battery was modified to make tasks operationally relevant. Initial evaluation of the ARES-I suggest that the interface change produced an assessment tool with user acceptance and performance data consistent with previous ARES research.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multi- Disciplinary University Research Initiative (MURI) program, P.A. Hancock, *Principal Investigator*, administered by the Army Research Office under Grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

REFERENCES

Benedetto, J., Harris, W.C., & Goernert, P.N. (1995). Assessing gender differences and norm data on a cognitive performance measure. *Proceedings of the Human Factors and Ergonomics Society, 39*, 968.

Baddeley, A. D. (1968). A three minute reasoning tested based on grammatical transformation. *Psychonomic Science, 10* (10), 341-347.

Elsmore, T.F., & Reeves, D.L. (2006). ANAM™ for palm OS handheld computers: The ARES test system for Palm OS handheld computers. *Neuropsychology Review,* in press.

Englund, C.E., Reeves, D.L., Shingledecker, C.A., Thorne, D.R., Wilson, K.P., & Hegge, F.W. (1987).*Unified Tri-service cognitive performance assessment battery (UTC-PAB). I. Design and specification of the battery.* Report No. 87-10, Naval Health Research Center, San Diego, CA.

Harris, W.C., & Hancock, P.A. (2003). Field assessment of cognitive performance under stress, *Proceedings of the Human Factors and Ergonomics Society, 47,* 1953-1957.

Harris, W.C., Hancock, P.A., & Harris, S.C. (2005.). Information processing changes following extended stress. *Military Psychology*, *17* (2), 115-128.

Harris, W.C., Hancock, P.A., & Morgan, C.A. (2005). Cognitive change following Special Forces training, *Proceedings of the Human Factors & Ergonomics Society, 49,* 1776-79.

Reeves, D.L., Winter, K.P, LaCour, S.J., Raynsfor, M., Vogel, K., & Grissett, J.D. (1991). *The UTC-PAB/AGARD stress battery: User's manual and system documentation*. (Special Report 91-3). Pensacola, FL: Navy Aerospace Medical Research Laboratory.

Sheridan, T.B. (1987). Supervisory control. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp.1243-1268). New York: Wiley.

Thorne, D.R., Genser, S.G., Sing, H.C., & Hegge, F.W. (1985). The Walter Reed Performance Assessment Battery. *Neurobehavioral Toxicity and Teratology, 7*, 415-418.

.

Application of Fuzzy Signal Detection Theory to Threat Detection in Extreme Environments

P.A. Hancock[1], J.L. Szalma[1], M.E. Koltko-Rivera[2], and D.A. Vincenzi[3]
[1] *University of Central Florida*
[2] *Professional Services Group*
[3] *NAVAIR Orlando*

e-mail: phancock@pegasus.cc.ucf.edu
Proposal Focus: Theoretical
Presentation Type: Poster
Non-student work

Abstract

Threat detection in extreme environments can induce considerable workload and stress on the individual due not only to the environment itself but the time compression that results from it (e.g., in high performance aircraft). Operators of complex systems in these environments must be able to detect critical signals quickly and accurately to avoid mission failure. Extreme environments are very unforgiving and mistakes can be fatal not only for the pilot but for many others in the affected realm. Detection systems and associated displays to support threat detection under stress must therefore be designed with these environments in mind. Here we discuss that application of Fuzzy Signal Detection Theory (FSDT), a recently developed extension of traditional Signal Detection Theory (SDT), to threat detection in extreme environments.

Summary

In the SDT model the state of the world is forced into crisp mutually exclusive categories (i.e., signal versus non-signal; old versus new; etc.), which may not be accurate representations of the true states of the world that face decision-makers. To capture the uncertainty inherent in real-world situations the representation of the stimulus dimension must be permitted to be continuous. A recent advance in SDT promises to enhance measurement of performance in these domains by capturing the continuous-nature of real-world stimuli and also by incorporating into the model the uncertainty inherent in such stimuli and in the decision-makers who respond to them. This development, Fuzzy Signal Detection Theory (FSDT; Parasuraman, Masalonis, & Hancock, 2000), combines traditional SDT with Fuzzy Set Theory to extend signal detection analysis beyond the traditional crisp, categorical decision-making model. FSDT allows for events to simultaneously be in more than one state (e.g. signal and non-signal), so that a particular instantiation of a stimulus (e.g. randomly drawn from a stimulus continuum) has properties of both ‘signal’ and ‘non-signal.’ This allows stimulus and response dimensions to be continuous rather than categorical. Consequently, FSDT can be employed in applied settings where the degree to which an event is a signal for detection may vary and decisions must be made under conditions of high uncertainty. Indeed, recent evidence indicates that FSDT provides a better description of the effects of

stimulus uncertainty on observers' responses, while the traditional model is more sensitive to manipulation of task difficulty (Murphy, Szalma, Hancock, & Stafford, 2003). In addition, the normality assumption of traditional SDT seems to extend to FSDT (Szalma, Oron-Gilad, Saxton, & Hancock, in press; Murphy, Szalma, & Hancock, 2004). Comparison of the fuzzy and traditional models has indicated that fuzzy false alarm rates are lower than those computed using traditional methods (Masalonis & Parasuraman, 2003), and that fuzzy ROCs indicate higher sensitivity than their crisp counterparts (Szalma et al., in press; Murphy et al., 2004). In addition, provision of multiple response categories results in more accurate signal detection in FSDT than forcing a binary response (Szalma et al., in press).

The Fuzzy Mapping Function

The key to successful application of FSDT in any domain is the establishment of an accurate mapping function. This function formally defines the fuzzy stimulus dimension, **s**, ranging from 0 to 1, in which 0 represents full membership in the 'non-signal' category and 1 represents full membership in the 'signal' category. In the mapping function **s** is related to a physical variable or set of variables, and thereby imposes meaning on the fuzzy stimulus and response dimensions. For instance, a fuzzy signal of **s**=.5, a point of maximal uncertainty along the fuzzy dimension, must be defined by a particular characteristic or set of characteristics of the real-world variable (e.g., a distance to target). An additional complication is that mapping functions can change according to environmental conditions. Thus, a physical variable that is important in one context is not important in another. A successful detection system must have these relations and their associated dynamics incorporated into their design.

Application of Sequential Dependency Analysis to Mapping Function Derivation

Behavior in the real world is not a single, discrete event. Rather, behavior occurs as a continuous stream in which events (and responses to them) at time (or event) *n* influences the behavioral response at time (or event) *n+1*. Current signal detection models, and the detection systems that are derived from them, are quite useful in describing the summative responses of detectors to large numbers of discrete events. However, because they relay on summated data these models cannot adequately capture the dynamics of detection; that is, how signals resolve over time. Combining traditional sequential dependency analysis (e.g., Treisman & Williams, 1984) with phase space representations of those dependencies will permit us to evaluate systematically how a response to an event, *n* (i.e., a correct identification, a false alarm, etc) influences a response to event *n+x*, where *x*=1, 2, 3… depending on the lags that show the most stable patterns and are most useful for prediction of performance in a given domain. Representing these in a phase space diagram provides a description of the operational characteristics of the detection system over time.

Future Research

Further work is necessary to investigate the relation between task difficulty and the FSDT model. In addition, there is evidence that the FSDT model captures the shifts in criterion setting resulting from changes in the distribution of 'signal-like' stimuli (Szalma, Stafford, & Hancock, 2003), similar to patterns observed in prior vigilance research in which changes in signal probability were evaluated using the traditional SDT model (e.g. see Parsons et al., 2000; Vickers & Leary, 1983). Future research will further examine the FSDT model by investigating the performance effects of variations in the distribution of stimulus uncertainty. In addition, further exploration of fuzzy ROC space is anticipated. Specifically, the conditions under which the equal variance assumption is met needs to be established, and the effect of stimulus and response range needs to be explored more thoroughly.

References

Masalonis, A.J., & Parasuraman, R. (2003). Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics, 46*, 1045-1074.

Murphy, L.L., Szalma, J.L., Hancock, P.A., & Stafford, S.C. (2003, March). *Comparison of methods of analyses from fuzzy and traditional signal detection theory*. Paper presented at the Mid-Year Symposium of Division 21: Applied Experimental and Engineering Psychology, Division 19: Military Psychology of the American Psychological Association and the Potomac Chapter of the Human Factors and Ergonomics Society, Ft. Belvoir, VA.

Murphy, L., Szalma, J.L., & Hancock, P.A. (2004). Comparison of fuzzy signal detection and traditional signal detection theory: Analysis of duration discrimination of brief light flashes. *Proceedings of the Human Factors and Ergonomics Society, 48*, 2494-2498.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42*, 636-659.

Parsons, K.S., Warm, J.S., Matthews, G., Dember, W.N., Galinsky, T.L., & Hitchcock, E.M. (2000, November). Changes in signal probability and vigilance performance. Poster presented at the meeting of the Psychonomic Society, New Orleans, LA.

Szalma, J.L., Oron-Gilad, T., Saxton, B., & Hancock, P.A. (in press). Application of fuzzy signal detection theory to the discrimination of morphed tank images. *Proceedings of the Human Factors and Ergonomics Society, 50*.

Szalma, J.L., Stafford, S.C., & Hancock, P.A. (2003). Application of fuzzy signal detection theory to vigilance: The effect of criterion shifts. Manuscript under review.

Treisman, M., & Williams, T.C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91*, 68-111.

Vickers, D., & Leary, J.N. (1983). Criterion control in signal detection. *Human Factors, 25*, 283-296.

ROAD ENVIRONMENT AND DRIVER FATIGUE

Tal Oron-Gilad, Peter A. Hancock
MIT^2 Laboratory
Department of Psychology & Institute for Simulation and Training
University of Central Florida
Orlando, Florida, USA
E-mail: torongil@mail.ucf.edu

Summary: We distinguish between fatigue caused by the demands of the driving task itself (see Hancock & Desmond, 2001) from the standard traditional approach that links fatigue predominately to the lack of sleep. Fatigue can be caused by two sources: (1) the driver's initial state before starting the drive, or (2) the characteristics of the drive and the road environment; both sources can have a cumulative effect. It is not clear what principles are involved in making one road environment more prone to inducing driver fatigue than another. For the purpose of the current presentation we provide empirical data on road environment and driver fatigue summarized from a series of three experiments that the first author has conducted at Ben-Gurion University (see Oron-Gilad, 2003; Oron-Gilad, et al., 2001). Those are examined in relation to the Hancock and Warm (1989) model of adaptability. The most significant and consistent findings of the three experiment is in the way that fatigue is reflected in driving performance across different road environments. These findings suggest that drivers are flexible in the way they handle fatigue over the course of time. They can adopt different strategies to compensate for their performance decrement, by focusing efforts on critical elements of each different type of roadway. Understanding of this dependency of fatigue symptoms on road conditions is of especial relevance to designers of technological fatigue countermeasures as well as those of future roadway systems.

INTRODUCTION

While the term driver fatigue is used frequently, it is a very general term and is used in many different ways. The one statement most agreed upon is that a fatigued driver is not a safe driver; however, the situations or the reasons for this impaired state may vary. We distinguish fatigue caused by the demands of the driving task itself (see Hancock & Desmond, 2001) from the more standard traditional approach that links fatigue to the driver's state and predominately to the lack of sleep (see Mackie & Wylie, 1992). Thus, fatigue can be caused by two sources: (1) the driver's initial state before starting the drive, or (2) the characteristics of the drive and the road environment. Both sources can have a cumulative effect. Regardless of the cause of fatigue, it is an unwanted state. However, different causes of fatigue may require different means of intervention. Specifically, fatigue caused by lack of sleep can be corrected by sleep, a strategy that may not improve the passive fatigue caused by a monotonous road. Furthermore, it is not clear what makes one road environment more prone to driver fatigue than others.

Resource theories and adaptive models of stress and sustained performance provide contrasting explanations for the effects of the driving-task demands. Resource theories (e.g., Wickens, 1984)

are frequently applied to performance of multiple component tasks such as driving. Those propose that an individual can be characterized by a limited supply of capacity of data both for attention and processing. A decrement in performance efficiency can be found when there is an insufficient quantity of resources unable to match the demands for resources made by the task or tasks to be performed. The actual performance is determined by the individual's strategy for allocating resources across the various tasks. For example, in order for an overloaded driver to maintain effective vehicle control, there may be a reduction in the efficiency of processing information from the traffic environment. Existing research provides a number of demonstrations of performance overload where task demands are increased by imposing additional task components, consistent with the prediction of resource theory (e.g., Brown, Tickner & Simmonds, 1969; Harms, 1986; 1991). On the basis of the same rationale, Hancock, Wulf, Thom and Fassnacht (1990) have shown that vehicle turn sequences are associated with greater attentional demands than straight driving. Multiple resource theories also propose that qualitatively different tasks may utilize independent reservoirs of capacity, such that dual task interference is reduced when tasks are dissimilar. Since driving is primarily a visual-motor task, the indication may be that in-vehicle countermeasures should utilize different input and output modalities. Empirical studies do not always support the resource theory (for example, Matthews, Sparks & Bygrave, 1996).

Dynamic models of stress and sustained performance (Hancock & Warm 1989; see Figure 1) are based on the notion of adaptation to task demands. The Hancock and Warm model suggests that it may be difficult to adapt to conditions of both under-load and over-load. According to this model, the individual is often able to compensate for dynamic variations in workload and

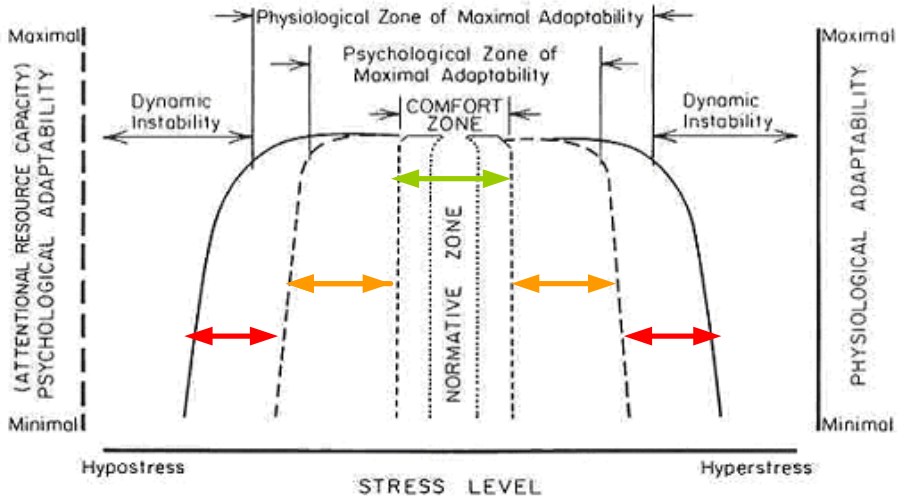


Figure 1. The extended-U model of stress and performance initially proposed by Hancock and Warm (1989)

environmental factors that moderate levels of stress. The potential hazard of fatigue is that matching effort to task demands may be impaired because fatigue reduces the range or efficiency of strategies available for regulation of effort. Adaptation to task demands is particularly relevant in driving since the task includes rapid variations in workload, as shown in Figure 2. Generally, under normal conditions, as shown in Figure 2(a), the contextual demands of the drive are lower than the driver's optimal attentional capacity. However, one of the great problems of powered

travel is that normal circumstances can rapidly change into conditions requiring emergency response. As shown in Figure 2(b) drivers respond to emergencies by focusing all of their attention onto the situation; the so called “narrowing of attention” effect (see Hancock & Weaver, 2005).

Paradoxically, when the task was relatively difficult (curved road), fatigued drivers were more often able to cope with increased demands; when it was easy (straight road) performance tended to deteriorate, implying that fatigued drivers are failing to adjust their effort effectively (Desmond & Matthews, 1997). Hence, complacency problems may increase as a result of fatigue in consistency with the Kahneman’s (1973) model that the resource availability is variable. More recent analyses (Hancock & Warm 1989; Hockey 1993; 1997) have supported Kahneman's conclusions, indicating other costs of performance deterioration, such as attentional errors and reduction in subsidiary aspects of task performance. The compensatory control model described in detail in Hockey (1993; 1997) suggests that the system is typically biased toward maintaining high-priority goals under stress at the expense of neglecting the other goals.

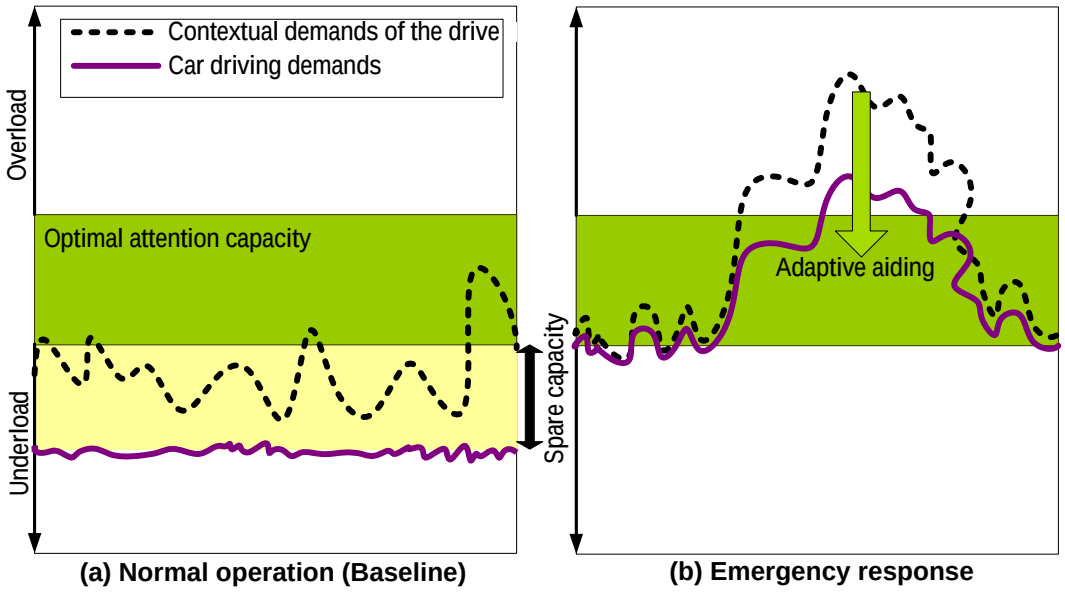


Figure 2. Attention capacity and driving demands for drivers

For the purpose of the current presentation, we provide empirical data on road environment and driver fatigue summarized from a series of three experiments that the first author conducted at Ben-Gurion University (see Oron-Gilad, 2003; Oron-Gilad, et al., 2001). Those are combined with theoretical implications developed by both authors in relation to the empirical data and to the Hancock and Warm (1989) model of adaptability. Difficulties in defining robust criteria for driver fatigue have been related mostly to the large individual differences among drivers (Brookhuis & de Waard, 2000). In a previous study (Oron-Gilad, et al., 2001), we suggested that difficulties in defining robust criteria for driver fatigue can also be traced to the use of different coping strategies in various road conditions. In Ronen, et al. (2002), we showed that the low demand (straight) road and the high demand (curved) road vary in the way they affect mental workload, which we measured by heart rate variability (HRV). In a prolonged drive we assume that the changing road situation will have an impact on driving performance measures as well.

We suggested that drivers adopt different fatigue-coping strategies relative to the demands and conditions of the task, hence, drivers may be aware of a general decrement in their performance due to fatigue, and the way they handle the decrement is related to the demands of the road.

METHODS

Three experiments examined the effect of road characteristics on fatigue and fatigue-coping strategies in simulated driving. For all three experiments, driving performance measures, reaction time measures, physiological (ECG and EEG) measures, and subjective measures were collected. A total of 27 drivers (10 professional military truck drivers and 17 students from Ben-Gurion University) participated in the experiments. Perceived fatigue related to the driving task was assessed by the Swedish Occupational Fatigue Inventory-20 (SOFI) (Aahsberg, 1998). The first experiment used three different road environments, each simulating a different road in the southern part of Israel; a winding road driven either uphill or downhill, a two-lane straight rural road, and a straight four-lane divided highway. The other two experiments used only the winding and the rural road environments. The experiments varied in the way the different roads were presented. In the first and second experiments, different road sections were combined into a single driving scenario, while in the third experiment, drivers drove either one of the two possible roads (winding or rural).

RESULTS

The analysis of results consistently showed significant differences in driving performance among the various road environments and changes over the course of time. Driving performance was measured by four parameters that were recorded by the simulator: the root mean square (RMS) of the lane position, RMS of the steering wheel rate, the average longitudinal speed and the RMS of the longitudinal speed. For the driving performance measures, the most striking results that were consistent across all three studies were that (1) there was a fatigue-related performance decrement, and (2) that decrement was manifested differently for each type of road. In each type of road, drivers chose to “loosen up” in what they perceived as the most tolerant element. On the two-lane winding road the change in performance appeared in the longitudinal speed, with speed being significantly higher as the driving progressed. On the two-lane undivided straight road, the deterioration in performance was significant for the quality of the lane positioning and the corresponding steering wheel control. There were no significant differences in speed. Figure 3 depicts the differences in steering wheel control in the third study. Even though these findings were consistent across all three studies, there were differences among the studies in the degree of change measured throughout the drive. Also, greater individual differences were identified among the professional military truck drivers than among student participants. As a group, the latter tended to drive more poorly from the initial period of the drive.

The subjective rating of fatigue, using the SOFI, indicated that after the completion of the drive, drivers felt a change in their sleepiness, in their motivation to continue driving, and in their level of energy, as well as an overall increase in feelings of fatigue.

The heart rate variability measure (derived from the ECG) provided additional evidence of the sustained performance effect of driving; time on task had a pronounced effect on drivers and fatigue onset varied across experiments but appeared after approximately 50 minutes. Figure 4

shows this phenomenon as it was captured in the third experiment, where participants drove either one of the roadways.

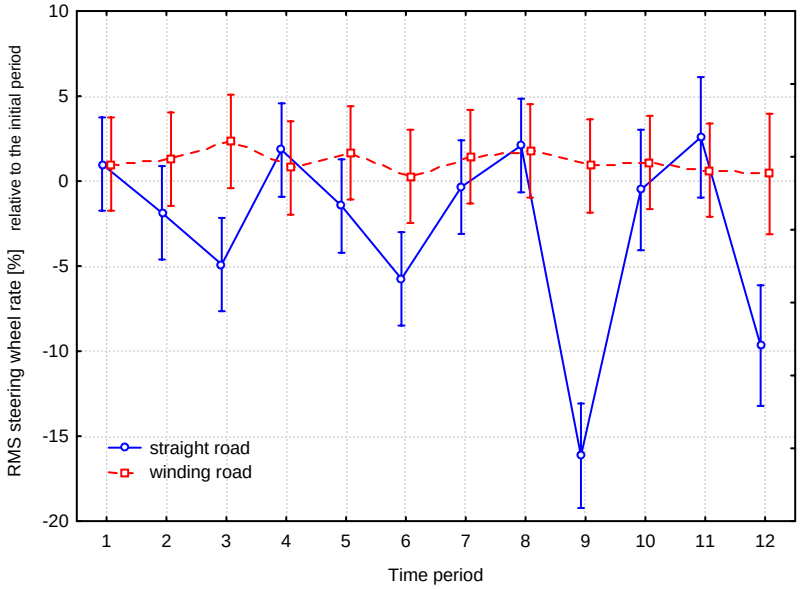


Figure 3. RMS steering wheel rate for the winding road and for the straight road relative to the initial period of the drive for driver in experiment 3 (the 60 minutes drive was broken into 12 time periods)

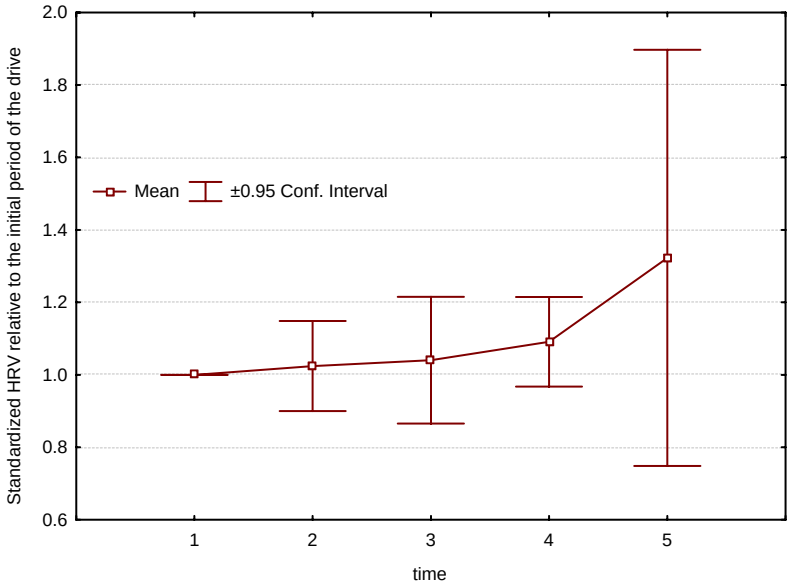


Figure 4. Standardized HRV relative to the initial period of the drive for all drivers in Experiment 3 (the 60 minutes drive was broken into 5 time periods)

DISCUSSION

The most significant and consistent findings of the three experiment is in the way that fatigue is reflected in driving performance across different road environments. Our results suggest that drivers are flexible in the way they handle fatigue over the course of time. They can adopt different strategies to compensate for their performance decrement, by focusing efforts on critical elements of the road. Understanding this dependency of fatigue symptoms on road conditions is of high relevance to designers of technological fatigue countermeasures. When driving performance decrement varies as the road conditions (or road demands) vary, one has to be very cautious in implementing countermeasures that are based on a single or on a limited number of performance measures. Theoretical models are therefore required to address this pattern and advance our understanding of driver behavior in relation to the road environment characteristics.

ACKNOWLEDGMENTS

This work was facilitated by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621, Peter A. Hancock, Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The data was collected when the first author was a Ph.D. student at Ben-Gurion University of the Negev, Beer-Sheva, Israel; David Shinar was the dissertation advisor and the study was supported in part by a grant from the GM foundation.

REFERENCES

Aahsberg, E. (1998). Perceived fatigue related to work. *Internal report, University of Stockholm, Department of Psychology, Sweden*. ISBN 91-7153-830-5.

Brookhuis, K.A., & de Waard, D. (2001). Assessment of Driver's Workload: Performance and Subjective and Physiological Indexes. In: Hancock P.A. and Desmond P.A. (Eds.), *Stress, Workload, and Fatigue* (pp. 321-333). London: LEA Publishers.

Brown, I.D., Tickner, A.H., & Simmonds, D.C.V. (1969). Interferences between concurrent tasks of driving and telephoning, *Journal of Applied Psychology*, *53*: 419-424.

Desmond, P.A., & Matthews, G. (1997). Implications of task induced fatigue effects for in-vehicle countermeasures to driver fatigue. *Accidents Analysis and Prevention, 29*(4): 513-523.

Hancock, P.A. & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*: 519-537.

Hancock, P.A., & Desmond, P.A. (Eds.). (2001). *Stress, workload and fatigue.* Mahwah, NJ: Lawrence Erlbaum.

Hancock, P.A., Wulf B., Thom, D., & Fassnacht, P. (1990). Driving workload during different driving maneuvers. *Accidents Analysis and Prevention, 22*: 281-290.

Hancock, P.A. & Weaver (2005). Temporal distortions under extreme stress. *Theoretical Issues in Ergonomic Science, 6*(2): 193-211.

Harms, L. (1986). Drivers' attentional response to environmental variations: A dual-task real traffic study. In: A. G. Gale, M. H. Freeman, C. M. Haslegrave, P. Smith and S. Taylor (Eds.), *Vision in Vehicles* I (pp. 131-138). Amsterdam: North Holland.

Harms, L. (1991). Variations in drivers' cognitive load. Effects of driving through village areas and rural junctions. *Ergonomics*, *34*: 151-160.

Hockey, G.R.J. (1993). Cognitive-energetic mechanisms in the management of work demands and occupational health. In: A. D. Baddeley and L. Weiskrantz (Eds.), *Attention, selection, awareness and control: a tribute to Donald Broadbent* (pp. 328-345). Oxford: Oxford University Press.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload. *Biological Psychology*, *45*: 73-39.

Kahneman, D. (1973). *Attention and effort*. Engle wood cliffs, NJ: Prentice Hall.

Mackie, R.R., & Wylie, C.D. (1992). Driver fatigue: Accidents, Assessments, and countermeasures. *Report no. 92-00169, Human Factors Research laboratory, Essex corporation*.

Matthews, G., Sparks, T. J., & Bygrave, H. M. (1996). Stress, attentional overload and simulated driving performance. *Human Performance, 9*(1): 77-101.

Oron-Gilad, T. (2003). *Alertness maintaining tasks while driving*. Ph.D. Thesis, Ben-Gurion University of the Negev, Beer-Sheva, Israel.

Oron-Gilad, T., Ronen, A., Shinar, D., & Cassuto, Y. (2001). Fatigue of professional Truck Drivers in Simulated Driving: A preliminary Study, Traffic Safety on Three Continents, International Conference in Moscow, Russia, 19-21 Sept.

Ronen, A., Oron-Gilad, T., Shinar, D., & Cassuto, Y. (2002). Assessment of driver's mental effort while driving roads of different complexity. In: de Waard D., Brookhuis K. A., Weikert C. M. and Toffetti A. (Eds.), *Human Factors in Transportation, Communication, Health, and the Workplace*. Maastricht, NL: Shaker Publishing.

Wickens, C.D. (1984). *Engineering psychology and human performance*. New York: Harper.

Running Head: Individual Differences Among Police Officers

Police Firearms Training: Individual Differences in Performance, Workload, and Stress

J.L. Szalma, T. Oron-Gilad, S.C. Stafford, and P.A. Hancock

University of Central Florida, Orlando FL

Address correspondence to James L. Szalma, Department of Psychology and Institute for Simulation and Training, University of Central Florida, 3100 Technology Parkway, Ste 333, Orlando, FL 32826-3281; jszalma@mail.ucf.edu

Keywords: Police; Performance; Workload; Stress; Coping; Personality

Abstract

We report what we believe to be the first study examining the relation between personality trait differences among police officers and their performance, workload, stress, and coping strategies within firearms operations tasks. In our study, police officers participated in four range-based shooting procedures of ascending task complexity. Personality traits predicted workload, stress, and coping responses. Surprisingly, Intellect/Openness to Experience emerged as the most dominant predictor among the 'Big Five' factors. In addition, there was evidence that trait interactions predicted performance and stress, indicating the need for future research to test not only for 'main effects' of individual traits but also for the interactions between these traits. These findings have significant implications for police officer selection and training.

Police Firearms Training: Individual Differences in Performance, Workload, and Stress

Contemporary concerns regarding homeland security and law enforcement indicate the need for better understanding of the factors that influence the performance of security personnel and police officers, including the personality characteristics of the individuals themselves (Hancock & Hart, 2002). The increased problems associated with law enforcement in the post-9/11 world impose considerable workload and stress, which can subsequently impair an officer's performance at critical moments. It is therefore crucial that we record and understand not only performance outcomes but also the workload and stress imposed by such work, and how law enforcement professionals cope with these demands. An important initial step is to explore how officer personality influences the relations among these variables. Some prior studies have examined the influence of personality on police performance and stress (e.g., see Alexander, Inees, Irving, Sinclair, & Walker, 1991; Brown & Campbell, 1990; 1994; see also Thompson, Kirk-Brown, & Brown, 2001). However, most of this research has focused on global measures. There have been relatively few studies examining these variables in the context of performance of specific but critical tasks police officers must perform. Indeed, even from an extensive literature search we have been unable to identify a single study that has examined the influence of police officers' trait differences on performance, workload, and stress on a *specific* work-related task. The purpose for the current study was to examine the relations among these variables. First we review briefly how the separate domains have been investigated and their implications for the performance of police officers.

Police Personality

Studies aimed at defining police personality profiles have produced diverse findings (see Brown & Campbell, 1994). When differences from normative values have been observed,

researchers have reported that police officers tend to show higher levels of psychological stability and mental health (Adlam, 1982; Biggam & Power, 1996; Carpenter & Raza, 1987; Gudjonsson & Adlam, 1983; Matarrazzo, Allen, Saslow, & Wiens, 1964; but see also Beutler, Nussbaum, & Meredith, 1988).

In regard to application of personality trait measures to the prediction of police performance, results have been mixed (Brown & Campbell, 1994; Burbeck & Furnham, 1985). One limitation noted by Burbeck and Furnham (1985), is that there have been variable criteria for 'good' or 'successful' police officers. However, meta-analytic work has established that police officers' traits, like those of individuals in other professions, are related to global measures of job performance (e.g., Barrick & Mount, 1991; Barrick, Mount, and Judge, 2001; Salgado, 1997). In a second order meta-analysis, Barrick, Mount, and Judge (2001) reported small positive correlations between trait and performance for Extraversion, Emotional Stability, Agreeableness, and a very small (ρ=.03) correlation for Openness to Experience (Intellect; see Goldberg, 1992). As for other occupations in their analysis, Conscientiousness emerged as the strongest predictor of global measures of police performance (ρ=.26). However, these studies used multiple performance criteria that include evaluations by supervisors and *global* performance records. There has been a distinct lack of data regarding the link between police personality and the performance, workload, and stress associated with a *specific* job-related task as opposed to overall job performance. There have, of course, been many studies on personality and specific task performance in populations outside law enforcement.

Personality and Performance

Although the link between personality and specific task performance has been neglected in the police domain, there is a large literature that has examined trait-performance links across

different kinds of task. Within the context of the Five Factor Model, most of these studies have examined Extraversion and Emotional Stability/Neuroticism (for reviews see Eysenck & Eysenck, 1985; Matthews, Davies, Westerman, & Stammers, 2000; Matthews, Deary, & Whiteman, 2003). In general, high Extraversion is associated with superior performance on tasks that impose high working memory demands; require divided attention, and in contexts that are arousing. Low Emotional Stability is sometimes associated with lower performance, although results are dependent upon contextual factors (see Matthews et al., 2000; 2003). When performance differences are observed they likely result from reduction in resources available for performance (Hancock & Warm, 1989; Humphreys & Revelle, 1984) or from reduction in the efficiency of information processing (Eysenck & Calvo, 1992). While Emotional Stability and Extraversion have been widely studied, less attention has been paid to how well the other three 'Big Five' factors to predict performance, although such efforts are anticipated (Matthews et al., 2003).

Much of the research on the relation between personality and performance has focused on single traits. However, it is theoretically possible and, in practical terms very likely, that traits interact in their prediction of performance. That is, a particular trait may have a small or negligible correlation with performance when considered singly, but have a substantive relation when considered in the context of an interaction with another trait variable. A recent example is given by Burke and Witt (2002) who examined the regression of supervisor performance ratings on the interactions between Openness (Intellect) and three other trait variables, Extraversion, Emotional Stability, and Agreeableness. They omitted Conscientiousness because of the emphasis in their study on social variables. Their data revealed a significant interaction between Intellect and Extraversion and between Intellect and Emotional Stability. Thus, performance

ratings were higher for introverted rather than for extraverted individuals at lower levels of Intellect/Openness. Further, for individuals low in Intellect/Openness, those who were also low in Emotional Stability received lower performance ratings relative to those high in Emotional Stability. Given the potential for trait interactions predicting performance and subjective state outcomes, in this study we therefore examined whether Intellect and Emotional Stability jointly predict measures of performance, workload, and stress state that are task specific and relatively non-social. We also included a test for an interaction of Intellect with Conscientiousness, since a similar pattern might be expected. Performance differences across levels of Conscientiousness might be larger for those who are low in Intellect. In addition, for reasons articulated below, we also examined two other potential interactions, Extraversion by Emotional Stability, and Conscientiousness by Intellect.

Personality and Workload

There have been a few studies examining the correlation of personality with subjective reports of mental workload (for an early review see Damos, 1988). These studies have examined specific variables that correlate with the five factors (e.g., cognitive complexity, Robertson, 1984; Robertson & Meshkati, 1985; Type A behavior patterns, Damos & Bloem, 1985; cognitive failure, Grubb, et al., 1995) rather than the factors themselves. In one exception, Rose, Murphy, Byard, and Nikzad (2002) reported that none of the five factors predicted global workload from a vigilance task, and among the subscales of the NASA Task Load Index, a measure of perceived workload (Hart & Staveland, 1988), the only significant relation was an increase in Frustration with higher levels of Neuroticism. However, Penley and Tomaka (2002), using public speaking as a stressful task, observed that perceptions of own-performance (a component of workload) were positively correlated with Extraversion and Conscientiousness, and negatively correlated

with Emotional Stability/Neuroticism. They also reported that appraisal of task demand was negatively correlated with Intellect and Conscientiousness.

Personality and Stress

Research on the relation between personality and self-reports of stress in the context of task performance has been examined by Matthews and his colleagues (Matthews et al., 1999; 2002). To assess stress as a multidimensional construct, Matthews and his colleagues developed the Dundee Stress State Questionnaire (DSSQ), a well-validated and reliable self-report measure of mental states (cognitive, motivation, and affective processes) associated with task performance. Three broad state dimensions have been identified (Matthews et al., 2002). Distress, reflected in scales of Tense Arousal, Hedonic Tone (general mood), and Confidence and Control, indicates an overload of processing capacity. Task Engagement, reflected by Energetic Arousal, Concentration, and Motivation, indicates commitment to effort toward task goals. Finally, Worry, reflected by Task-Related and Task-Irrelevant Cognitive Interference, Self-Focused Attention, and low Self-Esteem, indicates self-evaluation. In the initial studies with the DSSQ, correlations between the 'big 3' state dimensions and the 'big five' traits were obtained, and were generally in the expected direction (Matthews et al., 1999). Thus, Conscientiousness correlated positively with task engagement, while worry correlated negatively with Emotional Stability and Conscientiousness. Distress correlated significantly with all five traits, with higher distress associated with higher scores on Neuroticism (lower Emotional Stability) and lower scores on the other four factors.

Using other subjective measures of stress, Penley and Tomaka (2002) observed patterns of stress responses generally consistent with those of Matthews et al. (1999; 2002). Thus, in their study Emotional Stability was associated with lower perceptions of stress and lower levels of

negative emotion (i.e., less Distress), Extraversion was associated with lower levels of stress and higher levels of positive emotion, and Openness and Conscientiousness were both associated with less perceived stress. Note, however, that the tasks used in the studies by Penley and Tomaka (2002) and Matthews and his colleagues (1999; 2002) were primarily information processing tasks conducted in laboratory or office environments. The trait-stress state relations associated with a real-world perceptual motor task such as shooting have not been investigated.

Personality and Coping

Data on coping with social stress (Penley and Tomaka, 2002) and stress arising from a rapid information processing task (Matthews & Campbell, 1998) indicate that the five factors influence coping strategies used by participants. Matthews and Campbell (1998) derived a coping inventory for task stress to assess three general coping strategies individuals use to cope with task demands (see Endler & Parker, 1990; Lazarus & Folkman, 1984): task focused coping, which are efforts to deal directly with the task demands; emotion-focused coping, which reflects efforts to regulate one's emotional response to the stress; and avoidant-coping, which diverts attention away from the task. They observed that both emotion-focused and avoidant coping were negatively correlated with Emotional Stability. Avoidant coping was also negatively correlated with Emotional Stability, Agreeableness, and Conscientiousness. For Emotional Stability, similar results have been observed in regard to general coping styles (e.g., see Kardum & Hudek-Knezevic, 1996). Penley and Tomaka (2002) observed that while low Emotional Stability related to use of more emotion-focused and avoidant coping type responses, they also observed that Extraversion, Openness/Intellect, and Conscientiousness were associated with coping responses consistent with problem or task focused coping. In sum, it appears that high levels of Extraversion, Intellect/Openness, and Conscientiousness are associated with the use of

adaptive coping strategies (i.e., more task- or problem-focused coping and less emotion-focused or avoidant coping), while low Emotional Stability, low Agreeableness, and low Conscientiousness are associated with less adaptive coping patterns.

Personality, Stress, and Time Perception

One of the ways in which individuals adapt to stress is to narrow the focus of their attention (Easterbrook, 1959; Hancock & Warm, 1989). However, such narrowing can induce distortions in perception that affect performance. While much of the work in this area has focused on the distortion of spatial perception (e.g., Easterbrook, 1959; Hancock & Dirkin, 1983), it has recently been argued that distortions resulting from attentional narrowing can occur along the temporal dimension as well (Hancock & Weaver, 2004). Further, Hancock and Weaver (2004) have argued that individual differences likely influence the distortions that occur. They noted, for instance, that under stress extraverts might exhibit increased attention to external cues while introverts would be more focused on internal cues. From this one would predict that Extraversion would be negatively related to underestimation of time (i.e., time would seem to 'slow down' for those high in Extraversion). Indeed, Zakay, Lombranz, and Kziniz (1984) observed that individuals high in Extraversion estimated intervals to be longer relative to individuals lower on this trait when the task was not cognitively complex (although see Kirkaldy, 1984; see also Rammsayer, 2002). In regard to Emotional Stability, given that individuals who are low in this trait have fewer cognitive resources to devote to task performance (e.g., see Eysenck & Calvo, 1992; Matthews et al., 2003); they might therefore be less able to effectively process environmental cues, thereby inducing underestimation of time intervals. We tested these hypotheses in this study by asking officers to make prospective time estimations while performing two relatively novel and challenging tasks (see methods section).

Resultant Hypotheses

The above considerations indicate that although the personality characteristics of police officers have been examined in relation to global measures of performance and stress, they have not been linked to performance on specific tasks. Examination of the relation between traits and the workload, stress, and coping associated with performing specific tasks has been examined, but not in relation to a perceptual motor task (shooting) performed by police officers. The current study was designed to address these gaps in the literature by an initial exploration of the associations among these variables. Indeed, we believe such studies are necessary to building the database to support development of a general theory of personality and human performance, a crucial undertaking since there is to date no such theory that is generally accepted (Matthews, Deary, & Whiteman, 2003; but see also McCrae & Costa, 1999). It is therefore our expectation that these data will inform theory development and further empirical efforts. In the following sections we articulate tentative hypotheses regarding how each of the five traits and specific interactions between traits, is expected to relate to performance, workload, stress, and coping response. The clearest predictions are in regard to Extraversion and Emotional Stability, given the larger research pedigree associated with these characteristics. Across all traits, however, the expected predictions of performance and workload were hypothesized to be related to two novel and challenging tasks performed by the officers rather than two relatively familiar and simpler tasks. That is, the higher skill level for the simpler tasks was expected to attenuate the trait differences, and the resource demands of the more challenging tasks should enhance trait differences. Summaries of the expected regressions for each trait are shown in Table 1.

Extraversion

Given that extraverted individuals tend to prefer arousing environments and tend to do

better than introverts on tasks that place demands on working memory, it was expected that Extraversion should predict better shooting accuracy. Based on Hancock and Weaver's (2004) argument that the tendency of extraverts to attend to external cues will induce a subjective experienced of time passing more slowly, it was expected that time estimates will be positively associated with Extraversion. Note that for the time estimation tasks it was expected that participants across all trait levels would generally underestimate durations, since previous research has shown that such effects are likely when time estimation is not the primary task at hand (Zakay, 1998; see also Block, Zakay, & Hancock, 1999; Zakay & Block, 2004). Allocating fewer attentional resources to temporal information processing causes fewer time signals to accumulate, resulting in a decrease in the estimated duration. Thus, higher Extraversion should be associated with less underestimation of time intervals.

If performance differences associated with different levels of Extraversion are related to superior working memory capacity and a preference for arousing conditions such as a shooting task, one might also observe lower levels of workload among those higher on that trait. Extraversion is also generally associated with positive affect, lower levels of stress (e.g., lower levels of distress and higher levels of task engagement) would be expected both prior to and after the session. Extraversion was expected to also predict task-focused coping, since in general they engage in problem-focused coping with life stress (McCrae & Costa, 1986). That is, individuals higher on Extraversion were expected to report greater use of adaptive coping strategies.

Emotional Stability

Previous research has established that the association between Emotional Stability and performance is often task dependent (see Matthews et al., 2003), and no data exist for the task

and population examined in the current study. However, when performance differences are observed, Emotional Stability is usually associated with better performance. Similarly, given the internal focus often associated with emotional instability, it was predicted that Emotional Stability would predict more accurate time estimates (i.e., a smaller degree of underestimation). In regard to subjective measures, individuals high in Emotional Stability should report lower perceived workload, and lower stress levels. Thus, Emotional Stability should predict lower distress and worry and greater task engagement both prior to and after task performance. Emotional Stability should be positively related to task-focused coping and negatively related to emotion-focused and avoidant coping. That is, Emotional Stability should be associated with more effective coping.

Conscientiousness

It was expected that Conscientiousness would predict better performance, particularly since in general this trait has been one of the strongest predictors of global performance measures (e.g., see Barrick, Mount, & Judge, 2001). However, the tendency for individuals high in Conscientiousness to work harder may also result in greater perceived workload. In regard to stress, those high in Conscientiousness should report higher task engagement and less distress both prior to and after the tasks (Matthews et al., 1999). Individuals high in Conscientiousness may report greater task-focused coping and less avoidant coping.

Agreeableness

Given the general finding of positive correlation between Agreeableness and performance (Barrick et al., 2001), it was expected that this relation would also occur in this context. No predictions were made in regard to perceived workload. Due to their higher degree of positive affect, cooperativeness, and tolerance for 'going with the flow,' it was expected that these

individuals would report lower levels of distress and worry and higher levels of task engagement both prior to and after the session. In regard to coping, Agreeableness was expected to be associated with greater task-focused coping (as part of their tendency to be co-operative) and less avoidant coping.

Intellect (Openness to Experience)

No predictions for shooting accuracy were made in regard to Intellect, given the small correlation observed in meta-analyses of global performance (e.g., see Barrick, Mount, & Judge, 2001). However, as described later, interactions involving Intellect were expected. It was also expected that those low in Intellect would be more likely to exhibit inaccurate time estimations, based on work of Rammsayer (2002) who reported more accurate time estimates performance by those high in Psychoticism. Although this trait is not equivalent to Intellect, one of the explanations offered by Rammsayer (2002) for his findings was the greater creativity among those high in Psychoticism. Creativity is also a facet of Intellect. Given the expectation of general underestimation, it is therefore likely that high Intellect would be associated with less underestimation. Indiviuals high in Intellect might also be expected to experience lower workload (Penley & Tomaka, 2002). In addition, higher Intellect scores may predict lower levels of stress, given that these individuals prefer interesting and challenging activities and are better at adapting to change (Burke & Witt, 2002; Goldberg, 1992; but see also Matthews et al., 1999). Intellect should be associated with greater task-focused coping and less avoidant coping (see Penley & Tomaka, 2002), although the null results reported by Matthews and Campbell (1998) render these predictions tentative.

Trait Interactions

Based on the findings of Burke and Witt (2002), it was predicted that although Intellect

by itself might not predict improved shooting accuracy, it might interact with Emotional Stability, with Intellect more likely to be associated with performance differences at low levels of Emotional Stability. Specifically, the combination of low Intellect (e.g., low creativity, low ability to adapt to change) and low Emotional Stability is likely to be linked to lower levels of performance. Similarly, it is also possible that Intellect might interact with Conscientiousness. Note that the latter trait is associated with the tendency for determination, attentiveness to detail and precision, and greater task focus and achievement striving. Thus, the negative performance and higher stress level expected among those low in Intellect might occur only for those who are also low in Conscientiousness. A third potential interaction is that of Extraversion with Emotional Stability. That is, performance differences as a function of Extraversion might be larger for those low in Emotional Stability.

In regard to time estimation, errors are likely to be larger for those low in Extraversion and Emotional Stability, since these individuals are more likely to experience negative affect and be inwardly focused (Neurotic Introverts; see Eysenck & Eysenck, 1985). It is also possible that Intellect would interact with Emotional Stability, such that high scores on Intellect may compensate for low Emotional Stability. That is, the larger errors expected from those low in Emotional Stability may be exacerbated by those who are also low in Intellect. Similarly, one might expect that the relation of Intellect to time estimation would emerge more strongly in those low in Conscientiousness, since high scores on the latter might compensate for low scores on the former. Thus, larger errors in time estimation might be exhibited for those low on both Intellect and Conscientiousness.

The combination of low Extraversion with low Emotional Stability was expected to result in individuals with less resource capacity for task performance. Workload would then be

expected to increase with decreased Extraversion, but this increment would be greater for those low in Emotional Stability. The combination of low Extraversion and low Emotional Stability was also expected to be related to greater distress and worry and less task engagement. Similarly, the combination of low Intellect and low Emotional stability, and low Intellect and low Conscientiousness, should each result in higher perceived workload and stress reported by these individuals.

The interaction predictions are summarized below:

- *Extraversion by Emotional Stability*: The differences in performance, workload, and stress across the Extraversion scale will be larger among individuals low in Emotional Stability.
- *Intellect by Emotional Stability*. The differences in performance, workload, and stress across the Emotional Stability scale will be larger among individuals low in Intellect.
- *Intellect by Conscientiousness*. The differences in performance, workload, and stress across the Conscientiousness scale will be larger among individuals low in Intellect.

Method

Participants

Participants in this study were police officers in a Department in the Southeastern United States. All officers in the Department were required to participate in the night shooting exercise as part of their continuing training. However, participation in the present research was voluntary and each individual officer decided whether or not to participate in the research while performing the obligatory exercise. Of the 91 officers who completed the exercise, 71 officers (78%) volunteered to participate. These included 10 women and 61 men, with a mean age of 37 years (range 22-56 years) and with an average of 11 years of police experience. Note that as a result of

missing data, the number of participants included in individual analyses ranges from 58-65.

Firearms

Each officer used a SIG (Schweizerische Industrie Gesellschaft) Sauer p226 9mm handgun. All officers used the same standard issue duty belt and equipment. Each officer also carried a standard Department issued Streamlight Stinger flashlight 7.4 inches long and weighing 10 oz which they used in three of the four tasks. The only task which they were not required to hold the flashlight while shooting was the Warm-up Task (WT; see below). All officers were required to wear sound attenuating hearing protection (external muffs or internal plugs), body armor, and clear eye protection during each shooting task.

Task Design

The night shooting exercise included four different shooting tasks that varied in complexity and in their settings. Those were: the 'Warm-up Task' (WT), the 'Basic Flashlight Task' (FT), the 'Barrel Task' (BT) and the 'Metal Task' (MT). The tasks increased in difficulty as the training session progressed, with the 'Barrel Task' and the 'Metal Task' being the more complex of the four. The tasks were designed by the training officers and the authors were not involved in the design or the structuring of the tasks. All officers were familiar with the Warm-up Task (WT) and with the use of a flashlight while shooting from previous training exercises. The 'Barrel Task' (BT) and the 'Metal Task' (MT) were introduced for the first time in this exercise and therefore were novel tasks for all participants.

Warm-up Task (WT) – 'Baseline' Shooting Skill. The first two tasks (WT and FT) were group exercises where teams of officers participated at the same time under the direction of the training officer. WT was a draw and shoot task in which each officer was required to fire a total of 16 rounds from 4 different distances (5, 15, 21, and 35 feet) respectively at a standard paper

target which presented a blue silhouette against a white background. All rounds were fired in very low ambient light conditions in which the silhouette of the target was just visible. All officers were operating under the command of the instructor and were instructed to draw their weapon and engage the target upon the verbal command 'gun,' but not to move while engaging the target. After firing the required number of shots, officers were required to keep cover on their assigned paper target until instructed to 'holster' their weapon by the instructor. Although the WT was treated as a baseline task, it was part of the officers' performance evaluation and those who performed below the level of instructor expectations were required to repeat the task on the following training day (for the purpose of this study, data was recorded for each of these officers only on the first day of training). Performance on this task was evaluated by the percentage of hits out of the sixteen shots fired. A hit was defined as a shot placement on the target that was consistent with a wound which was expected to kill or incapacitate an opponent; a miss was defined as a shot placement which outside the silhouette target or that was inconsistent with a mortal or incapacitating wound.

Flashlight Task (FT) – Basic Flashlight Control and Movement Exercise. FT was a twenty-four round shooting exercise with reloads, while holding the flashlight and moving (i.e., rotating the body position to face 90 degrees or 180 degrees away from the target). Some portions of the task required officers to fire their weapons in very low ambient illumination conditions, whereas at other points officers were required to illuminate targets with their flashlights. In addition to shooting with a flashlight officers had to load their weapons with two plastic 'dummy' rounds placed into two of their three magazines. The dummy rounds imposed an additional task demand by jamming the weapon and requiring the officer to clear the weapon of the dummy round and continue to engage the target with appropriate number of rounds. All

rounds in this task were fired from a fixed distance of 21 feet. Performance on this task was evaluated by the percentage of hits out of the 24 shots fired. A hit was defined as a shot placement that was consistent with a wound which was expected to kill or incapacitate an opponent; a miss was defined as a shot placement which missed the silhouette of the target or was inconsistent with a mortal or incapacitating wound.

The 'Barrel' Task (BT) – Movement and Taking Cover while Shooting Exercise. Unlike the first two tasks, which were group exercises, this task was performed by each officer separately. The officer was required to sequentially shoot at five targets, arrayed horizontally and equally distributed over approximately 90 feet. At the beginning of the task the officers faced away from the first of the 5 targets with a loaded weapon. On the command of the instructor he or she engaged the five targets at different ranges with a specific number of rounds per target. After engaging each target, the officer ran to the next shooting location to engage the next target. On targets two and four, the officer took cover behind a plastic barrel and illuminated the target with a flashlight prior to engaging it. Performance on this task was evaluated by the percentage of hits out of the 15 shots fired. In addition, percentage of hits was calculated for each one of the 5 targets separately.

The 'Metal' Task (MT) – Feedback Shooting Exercise. Like the BT, this task was performed separately by each officer. The task was conducted in an area illuminated by the strobes of a police cruiser. This was a timed task where officers were required to complete the task as quickly as possible. Each officer had three magazines with six rounds in each. At the beginning of this task the officer was standing with his/her back against a wooden wall in a grassy area of the range. The wall was approximately 15 feet wide and had a 2 foot by 3 foot opening in its center. When the timer sounded with a loud audible beep the officer had to draw

both flashlight (to illuminate the target) and gun and turn 180 degrees and look through the opening in the center of the wall. Twenty-three feet from the officer were 6 metal targets. Each metal target was posted on a 3 feet and 8 inches pole with a 12 inch diameter target centered at 26 inches from the ground. The targets were separated by a 6 foot space into 2 groups of three. Within the 2 groups the targets were 16 inches apart. Each group of three included a white target, a red target, and a blue target. When the officer pointed the weapon toward the targets, the range instructor called out a color. The officer was then required to shoot each metal target in that color for 2 times consecutively (4 shots altogether). Once completed, the instructor would instruct the officer to proceed to the second part of the task. The officer had to move around the wooden wall through the 6 foot gap of the first set of metal targets and up to a the second firing line (a total movement of approximately 45 feet). The second firing line was 30 feet away from 8 metal targets separated into 2 groups of 4. Each target had a center mass measuring 12 inches by 12 inches with a 6 inch by 6 inch head area on top. The bottom of the targets was located 39 inches above the ground. Each group of four targets had a white target, black target, yellow target, and red target. The groups of targets were separated by a 3 foot space. Within each group targets were 16 inches apart. When the officer reached the second firing line the instructor would call out a color. The officer then had to hit each target with that color with 2 consecutive rounds. The timer was stopped when either the officer expended all rounds or had hit all metal targets. In sum, officers were required to hit two targets twice at each distance (23 and 30 feet accordingly), a total of 8 hits for all 4 targets. Performance on this task was evaluated by the percentage of hits out of the total shots fired (note that in this task the number of hits was constant (8) while the number of possible shots varied among officers).

Contrary to the BT, the MT included feedback after each shot as a result of the sound of

the bullet hitting the metal target. Note that although the participants were wearing hearing protection, it provided attenuation of the noise but did not completely mask the sound of the bullet hitting each target. Also, the targets differed in shape from the paper targets used in the other three tasks and were novel to most officers who usually train by shooting at the paper silhouette targets. Similar to the BT, the MT tasks required the officer to run from target to target and to aim at targets following the instructions of the shooting range instructor. However, in the MT officers were required to aim at one of the three (for the shorter distance) or four (for the longer distance) possible targets which varied by color. In contrast, in the paper target tasks, there was only one target on each trial (in the BT officers shot 5 different targets consecutively).

Performance Assessment

There are no State enforced rules regarding the criteria for passing the shooting exercise. The State requirements specifically address shooting performance during the day in a stationary position. Performance levels on the night training are therefore left to the judgment of the training officer. Since the BT and the MT were novel to the officers the training officer was not likely to pass or fail officers based on performance on these tasks. However, individuals who the training officers believed did not perform adequately on any one of the tasks was required to return to the shooting range on successive days until the training officers were satisfied with their level of performance. In these cases, data for this study was collected only on the first day.

Shooting performance was assessed separately for each task by calculating the percentage of hits from the total shots made. Time estimates (after the BT and MT) were made under the prospective paradigm, in which the officers knew in advance that they will be asked to judge the duration of the task (Block & Zakay, 1997). Clock time and estimated time scores were used to compute the duration judgment ratio (DJR; Block, Zakay, & Hancock, 1999), which is the ratio

of the estimated time and the clock time in minutes and seconds, expressed as a proportion.

Workload Assessment

Mental workload was assessed using a revised version of the NASA Task Load Index (RTLX). This instrument is used to assess perceived workload along six dimensions, three of which measure an individual's appraisal of task demands (mental, physical, and temporal demand), and three of which measure the individual's response to that demand (effort, ratings of own-performance, and frustration). The RTLX differs from the original NASA-TLX (Hart & Staveland, 1988) in that the un-weighted average of the subscale values is used as an index of global workload (see for example Fairclough, 1991). In the original version of the TLX, paired comparisons were used to derive weights for the six subscales of the TLX. However, as Byers, Bittner, & Hill (1989) have shown and as discussed in length in Nygren (1991), RTLX scores can provide a better account of the workload experienced by the participant than traditional weighted TLX scores. Following each task, officers rated their perceived workload on each of the six individual sources on a scale from 0-100. Note that in the case of own-performance ratings, the scores were reversed so that higher values indicated higher workload. Global RTLX estimates were derived by the calculating the unweighted average of the subscale values.

Stress State

Stress state was measured using the DSSQ, which was administered prior to performing any of the tasks and at the end of the training session. Unfortunately, the constraints imposed by the field training environment prevented collection of DSSQ data after each task. Standard scores on the DSSQ scales were computed by using the formula $\frac{(X_{Phase} - \overline{X}_{normative})}{s_{normative}}$

Normative means and standard deviations were obtained from Matthews et al. (1999).

Personality Traits

Traits in this study were measured using Goldberg's (1992) unipolar adjectival markers for the Big Five personality traits. This instrument consists of 100 adjectives, and each participant was instructed to rate the degree to which each adjective described him/her on a 9-point scale. Each of the five factors was derived from ratings for 20 adjectives. Note that in this measure, Neuroticism is labeled 'Emotional Stability' and Openness to Experience is labeled as 'Intellect.'

Experimental Procedure

The exercise occurred at an outdoor police shooting range in the Southeastern part of the United States. The average temperature at this period of time was approximately 50 degrees Fahrenheit (10 degrees Celsius) with no rain. The exercise was performed in the dark (after 7PM), and the total session time varied from approximately 60 minutes to 120 minutes depending on the number of officers in the session. Officers were trained in groups of 6 to 12 at a time. Prior to the beginning of the exercise the officers were debriefed by the training officer on safety precautions and on the tasks. The debriefing was held in a well lit lecture room which is part of the shooting range facility. At the end of the debriefing the officers were briefed on the research and its purpose and then asked to volunteer to participate. Informed consent was obtained from those who agreed to participate in the research. The experimenter then presented the questionnaires to be completed and briefly explained their components. Officers completed the adjective marker questionnaire followed by the pre-task version of the DSSQ. Officers were asked to complete the RTLX at the end of each task. The forms were distributed and completed in a lighted area behind the shooting range where the officers loaded their magazines throughout the exercise. Additionally, the officers were told that they would be asked by the experimenter

to provide an estimate of the task length (prospective duration estimation) in minutes and seconds upon completion of the BT and the MT. The post-DSSQ was administered after completion of all four tasks comprising the training exercise.

Results

Means and standard deviations for the trait measures and the DSSQ are displayed in Table 2. Descriptive statistics for the performance data and the time estimation task for the four tasks are shown in Table 3, and the RTLX data are shown in Table 4. The hypothesized relations among variables were tested via hierarchical regression.

Extraversion

Contrary to expectation, Extraversion did not predict performance on any of the shooting or time estimation tasks. In regard to workload, a marginal regression of temporal demand on Extraversion was observed for the FT task, $F(1,57) = 3.797$, $p = .056$, $R^2 = .062$; b = .349, SE_b = .179, β = .250, $t(57) = 1.948$, $p = .056$. All other regressions of workload scales on Extraversion were not statistically significant ($p>.05$ in each case). However, Extraversion significantly predicted greater pre-task engagement, $F(1,58) = 7.945$, $p = .007$, $R^2 = .120$; b = .013, SE_b = .005, β = .347, $t(58) = 2.819$, $p = .007$, and lower pre-task distress, $F(1,58) = 4.048$, $p = .049$, R^2 = .065; b = -.011, SE_b = .005, β = -.255, $t(58) = -2.012$ $p = .049$. Extraversion did not significantly predict post-task states ($p>.05$ in each case). Note that for this and all subsequent regressions involving the DSSQ, the post-task state was analyzed both with and without the pre-task state entered into the model first. This procedure established whether any associations between trait and post-task state were significant after accounting for pre-task state in the regression model. In regard to coping, no significant regressions of coping strategy on Extraversion were observed ($p>.05$ in each case).

Emotional Stability

Contrary to prediction, Emotional Stability did not predict performance outcomes (p>.05 in each case). However, Emotional Stability did predict scores on two workload scales, temporal demand and own-performance ratings. Thus, a significant regression of temporal demand on Emotional Stability in the BT task was observed, $F(1,57) = 4.251$, $p = .044$, $R^2 = .069$; b = .406, $SE_b = .197$, $\beta = .263$, $t(57) = 2.062$, $p = .044$. A similar but marginal regression was observed for the MT task, $F(1,57) = 3.588$, $p = .063$, $R^2 = .059$; b = .415, $SE_b = .219$, $\beta = .243$, $t(57) = 1.894$, p=.063. Emotional Stability was also related to own-performance ratings, with a significant regression for the MT task, $F(1,57) = 9.322$, $p = .003$, $R^2 = .141$; b = .533, $SE_b = .175$, $\beta = .375$, $t(57) = 3.053$, $p = .003$. Note that these trends were in the direction opposite to that predicted. That is, higher Emotional Stability predicted higher workload.

Higher levels of Emotional Stability predicted lower pre-task distress, $F(1,57) = 5.436$, $p = .023$, $R^2 = .087$; b = -.016, $SE_b = .007$, $\beta = -.295$, $t(57) = -2.322$, $p = .023$. The regression for pre-task engagement was not observed, however. A non-significant but marginal trend for the regression of pre-task worry on Emotional Stability was observed, $F(1,57) = 3.744$, $p = .058$, $R^2 = .062$; b = -.014, $SE_b = .007$, $\beta = -.248$, $t(57) = -1.935$, $p = .058$. The regression of post-task distress on Emotional Stability was significant, $F(1,57) = 4.720$, $p = .034$, $R^2 = .076$; b = -.023, $SE_b = .011$, $\beta = -.277$, $t(57) = -2.172$, $p = .034$. However, the increment in variance accounted for was not significant when the pre-state was entered first, suggesting that any association between this trait and post-task distress may be mediated by pre-task state. The regressions of post-task engagement and post-task worry on Emotional Stability were not statistically significant (p>.05 in each case).

As expected, an inverse relation between Emotional Stability and emotion-focused

coping was observed, $F(1,57) = 5.422$, $p = .023$, $R^2 = .087$; b = -.076, $SE_b = .033$, $\beta = -.295$, $t(57) = -2.329$, $p = .023$. More emotionally stable individuals engaged less in emotion-focused coping. Regressions of task-focused and avoidant coping on Emotional Stability were not statistically significant ($p>.05$ in each case).

Conscientiousness

Conscientiousness did not significantly predict performance on the shooting tasks or in time estimation ($p>.05$ in each case). In regard to perceived workload, a significant regression of temporal demand on Conscientiousness was observed for the FT task, $F(1,55) = 4.915$, $p = .031$, $R^2 = .082$; b = .435, $SE_b = .196$, $\beta = .286$, $t(55) = 2.217$, $p = .031$. Note, however, that contrary to prediction this regression was for one of the easier tasks rather than the more difficult tasks. A marginally significant regression of effort on Conscientiousness was observed for the MT task, $F(1,56) = 3.685$, $p = .060$, $R^2 = .062$; b = .292, $SE_b = .152$, $\beta = .248$, $t(56) = 1.920$, $p = .060$. This trend was consistent with expectation.

Higher Conscientiousness predicted lower pre-task distress, $F(1,56) = 3.996$, $p = .05$, $R^2 = .07$; b = -.012, $SE_b = .006$, $\beta = -.258$, $t(56) = -1.999$ $p = .05$. A non-significant trend in the expected direction was observed for the regression of pre-task engagement on Conscientiousness, $F(1,56) = 3.166$, $p = .081$, $R^2 = .054$; b = .010, $SE_b = .005$, $\beta = .231$, $t(56) = 1.779$, $p = .081$. Contrary to expectation, Conscientiousness did not predict post task state. However, a marginally significant regression of post-task distress on Conscientiousness was observed when pre-task distress was entered first in the regression, $F(2,55) = 9.311$, $p<.001$, $R^2 = .253$, $\Delta R^2 = .051$. For pre-task distress b = .826, $SE_b = .195$, $\beta = .510$, $t(55) = 4.227$, $p<.001$; for Conscientiousness b = .018, $SE_b = .009$, $\beta = .233$, $t(55) = 1.929$, $p = .059$. This indicates a trend for pre-task distress acting as a suppressor variable for the regression of post-task distress on

Conscientiousness. In regard to coping, no significant regressions on Conscientiousness were observed (p>.05 in each case).

Agreeableness

Agreeableness did not significantly predict performance, perceived workload, or pre-task stress state (p>.05 in each case). However, the regression of post-task engagement on Agreeableness was significant, and remained so after the pre-state was accounted for in the model, $F(2, 58) = 11.641$, $p<.001$, $R^2 = .49$, $\Delta R^2 = .05$. For pre task engagement, b = .794, SE_b = .12, β = .645, $t(58) = 6.873$, $p<.001$; for Agreeableness, b = .010 SE_b = .004, β = .225, $t(58)$ = 2.399, p= .02. Higher Agreeableness also predicted higher levels of task-focused coping, $F(1,59) = 5.740$, $p = .02$, $R^2 = .089$; b = .087, SE_b = .036, β = .298, $t(59) = 2.396$, $p = .02$. The regressions of emotion-focused and avoidant coping on Agreeableness were not statistically significant (p>.05 in each case).

Intellect

Intellect did not significantly predict shooting accuracy or time estimation (p>.05 in each case). In regard to perceived workload, there was an expectation that higher Intellect scores would predict lower workload. However, this trait predicted *higher* workload across the WT, FT, and MT tasks. Thus, for the MT task higher Intellect scores predicted increased global workload, $F(1,58) = 5.124$, p= .027, $R^2 = .065$; b = .258, SE_b = .114, β = .285, $t(58) = 2.264$, and mental demand, $F(1,59) = 6.714$, $p = .012$, $R^2 = .102$; b = .391, SE_b = .151, β = .320, $t(59)$ = 2.591, $p = .012$, and effort, $F(1,59) = 7.105$, $p = .01$, $R^2 = .107$; b = .351, SE_b = .132, β = .328, $t(59) = 2.665$, $p = .012$. For the FT task significant regressions were observed for global workload, $F(1,58) = 6.415$, $p = .014$, $R^2 = .100$; b = .256, SE_b = .101, β = .316, $t(58) = 2.533$, p = .014, mental demand, $F(1,58) = 4.587$, $p = .036$, $R^2 = .057$; b = .321, SE_b = .150, β = .271, $t(58)$ =

2.142, $p = .036$, and temporal demand, $F(1,58) = 6.258$, $p = .015$, $R^2 = .097$; b = .399, $SE_b = .160$, $\beta = .312$, $t(58) = 2.502$. $p = .015$. For the WT task there was a significant regression for performance workload, $F(1,56) = 4.721$, $p = .034$, $R^2 = .078$; b = .340, $SE_b = .156$, $\beta = .279$, $t(56) = 2.173$, $p = .034$. Thus, increased Intellect scores were associated with greater performance workload.

It was anticipated that higher Intellect would predict lower stress scores. Consistent with this expectation, significant regressions were found for pre-task engagement and pre-task distress. Higher Intellect scores predicted higher levels of pre-task engagement, $F(1,59) = 6.850$, $p = .011$, $R^2 = .104$; b = .011, $SE_b = .004$, $\beta = .323$, $t(59) = 2.617$, $p = .011$, and lower levels of pre-task distress, $F(1,59) = 4.621$, $p = .036$, $R^2 = .057$; b = -.010, $SE_b = .005$, $\beta = -.269$, $t(59) = -2.150$, $p = .036$. Although the regression of post-task distress on Intellect was not observed, the regression of post-task engagement on Intellect was statistically significant, and remained so after the pre-state was accounted for in the model, $F(2, 58) = 28.812$, $p<.001$, $R^2 = .498$, $\Delta R^2 = .06$, $F_{\Delta R}^2(1,58) = 6.96$, $p = .011$. For pre-task engagement: b = .702, $SE_b = .119$, $\beta = .578$, $t(58) = 5.886$, $p<.001$; for Intellect: b = .011, $SE_b = .004$, $\beta = .259$, $t(58) = 2.639$, $p<.011$. The regressions of pre- and post-task worry on Intellect were not statistically significant ($p>.05$ in each case).

Consistent with expectation, higher Intellect was associated with higher task focused coping, $F(1,59) = 6.461$, $p = .014$, $R^2 = .099$; b = .084, $SE_b = .033$, $\beta = .314$, $t(59) = 2.542$, $p = .014$, and with less emotion-focused coping, $F(1,59) = 4.198$, $p = .045$, $R^2 = .066$; b = -.049, $SE_b = .024$, $\beta = -.258$, $t(59) = -2.049$, $p = .045$. The avoidant coping regression was marginally significant, $F(1,59) = 3.85$, $p = .054$, $R^2 = .061$, b = -.044, $SE_b = .023$, $\beta = -.247$, $t(59) = -1.962$, $p = .054$.

Trait Interactions.

To facilitate analyses of trait interactions, the data for the predictor variables (each trait and an interaction term) were mean-centered prior to analysis (see Jaccard & Turrisi, 2003). Since these analyses were computed specifically to test for interactions between variables, only the test for the increase in variance accounted for by the addition of the interaction term was examined. In hierarchical regression this is equivalent to the test of the regression coefficient associated with the interaction term (Pedhazur, 1997).

Extraversion by Emotional Stability Interactions. A significant increment in variance accounted for by the interaction of Extraversion and Emotional Stability was observed for time estimation in the BT task, $t(54) = -2.123$, $p = .038$, $b_{IxC} = -.001$, $SE_b = 2.958 \times 10^{-4}$, $\beta = -.279$, $R^2 = .10$, $\Delta R^2 = .08$. The regression lines for time estimation (duration judgment ratio) as a function of Emotional Stability at three levels of Extraversion are shown in Figure 1. For this and subsequent graphs displaying interactions, 'Low' refers to one standard deviation below the mean on the respective trait and 'High' refers to one standard deviation above the mean. Note that these figures, and the division of scores into 'high,' 'low,' and 'average' levels on the trait, are intended only to facilitate understanding of the interaction. The formal test of interaction is in the test of the appropriate regression weight (Jaccard & Turrisi, 2003). It can be seen in the figure that the errors in time estimation associated with low Emotional Stability varies as a function of Extraversion. However, contrary to prediction, those high in Extraversion were *less* accurate in their estimates than those low in Extraversion. Differences among levels of Extraversion, however, were attenuated at higher levels of Emotional Stability, consistent with expectation.

The anticipated interaction of Extraversion and Emotional Stability in predicting

perceived workload was not observed (p>.05). However, the regression of post-task distress on the interaction between Extraversion and Emotional Stability was significant, t(53) = -2.137, p= .037, b_{IxC} = -.001, SE_b = 4.549 x10^{-4}, β = -.246, R^2 = .32, ΔR^2 = .06. Note that the pre-state was accounted for in the regression prior to entry of the interaction term. The ΔR^2 therefore represents the increment in variance in post-task distress accounted for by the interaction term after the pre-state has been accounted for in the model. The regression lines for post-task distress as a function of Emotional Stability at three levels of Extraversion are shown in Figure 2. For individuals low in Emotional Stability, those high in Extraversion experienced greater post-task distress than those low in Extraversion.

Interactions involving Intellect. The regression of MT performance accuracy on the interaction of Intellect and Conscientiousness was significant, t(53) = 2.455, p = .017, b_{IxC} = .012, SE_b = .005, β = 3.621, R^2 = .13, ΔR^2 = .10. The regression lines for MT accuracy as a function of Intellect at three levels of Conscientiousness are shown in Figure 3. It can be seen in the figure that the prediction of accuracy on the MT by Intellect depended on the level of Conscientiousness. Thus, individuals high in Conscientiousness performed better than those low on that trait, but only for those who were also high on the Intellect scale.

A significant increment in variance accounted for by the interaction between Intellect and Emotional Stability for time estimation associated with the BT task was observed, t(54) = -2.432, p = .018, b_{IxC} = -.001, SE_b = 2.985x10^{-4}, β = .312, R^2 = .14, ΔR^2 = .09. The regression lines for time estimation (duration judgment ratio) as a function of Intellect at three levels of Emotional Stability are shown in Figure 4. Consistent with expectation, differences in the accuracy of time estimation across levels of Emotional Stability emerged for individuals low in Intellect. For these participants, time estimates were less accurate for individuals who were also low on Emotional

Stability relative to those high on that trait. All other predicted regressions of performance on interaction terms involving Intellect were not statistically significant (p>.05 in each case). In addition, the regressions of perceived workload on the interaction between Intellect and Emotional Stability, and Intellect and Conscientiousness were not statistically significant (p>.05 in each case). However, a significant regression of post-task distress on the interaction between Intellect and Conscientiousness was observed, $t(52) = -2.105$, $p = .04$, $b_{IxC} = -.001$, $SE_b = 4.1639 \text{ x}10^{-4}$, $\beta = -.301$, $R^2 = .313$, $\Delta R^2 = .059$. The regression weight for the interaction term remained significant after pre-task distress was entered first in the model. The regression lines for post-task distress as a function of Intellect at three levels of Conscientiousness are shown in Figure 5. Thus, among those low in Intellect, individuals high in Conscientiousness reported greater post-task distress than individuals low in Conscientiousness.

A significant interaction between Intellect and Emotional Stability was observed for pre-task worry, $t(54) = 3.039$, $p = .004$, $b_{IxC} = .001$, $SE_b = 3.555 \text{ x}10^{-4}$, $\beta = .379$, $R^2 = .189$, $\Delta R^2 = .139$. The regression lines for pre task worry as a function of Intellect at three levels of Emotional Stability are shown in Figure 6. In this case, differences in worry across the levels of Emotional Stability emerged for individuals low in Intellect. Among these individuals, those low in Emotional Stability reported higher levels of worry than individuals high in Emotional Stability. A similar regression was observed for post-task worry, but the regression was not significant after the pre-state was accounted for in the model.

Discussion

The goal for the present study was an exploration of the influence of personality traits on the performance, workload, stress, and coping of police officers engaged in a shooting task. A summary of the outcomes is shown in Table 5. In regard to performance, it was expected that

personality traits would predict performance on tasks that were relatively novel and demanding. However, significant regressions were not observed for each individual trait. Apparently, even the novel tasks were similar enough to previous training or field experiences that the officers were able to perform effectively regardless of their level on a given trait. That is, the absence of direct associations between traits and performance may reflect the shooting skills of the officers who participated in this study. Their shooting skills were sufficiently well developed that they were able to adapt to the imposition of novel task demands (Hancock & Warm, 1989; Hockey, 1997). Indeed, that shooting skill is independent of personality among police officers is a desirable outcome for the safety of both the officers and the general public they serve. In regard to workload, stress, and coping, we address the findings separately for each trait. We then consider the trait interactions.

Extraversion

As expected, Extraversion predicted reduced pre-task stress, specifically greater task engagement and lower distress. This may reflect the generally positive affect associated with individuals high in Extraversion. The expected associations between Extraversion and perceived workload and coping strategies were not observed, indicating that findings reported previously may not extend to this population or to perceptual-motor tasks of the type examined in this study.

Emotional Stability

Emotional Stability predicted workload, specifically temporal demand in the BT task and own-performance ratings in the MT task. However, the direction of the predicted effect was opposite to that expected. Higher Emotional Stability was associated with *greater* perceived workload on these tasks, specifically the temporal demand and own-performance measures. Note that the temporal demand subscale of the TLX reflects appraisals of task demand, while own-

performance reflects individuals' appraisals of their response to that demand. It may be that emotionally stable individuals have more processing resources to devote to temporal components of the task and to monitoring their own performance, and thereby experience greater demand. The anxiety associated with low emotional stability may render these individuals less sensitive to workload. However, the relation of Emotional Stability to workload was task and workload scale specific. Emotional Stability predicted temporal demand in the BT task and (marginally) in the MT task, but this trait predicted own-performance ratings only in the MT task. The findings in regard to own-performance may have been due to a critical difference between the BT and MT tasks. Because the targets in the MT task were metal, the sound of the bullet hitting the target provided feedback to the officer regarding their performance, thereby making that aspect of perceived workload more salient. By contrast, no such immediate feedback was available for the other novel task, the BT. In this case the temporal demand of this timed task was likely more salient to the officers. Note that the dependency of the Emotional Stability-workload relation on the characteristics of the task extends to perceived workload the general finding that the influence of Emotional Stability on performance depends on context (Matthews et al., 2000; 2003).

In general, the expected associations between Emotional Stability and stress and coping were observed. Thus, higher Emotional Stability predicted less pre-task distress and less emotion-focused coping, confirming that results reported previously (Matthews et al., 1999; Matthews & Campbell, 1998) extend to law enforcement populations. Taken together with the absence of performance effects, these results suggest that although individuals low in Emotional Stability can perform the shooting tasks as accurately as those higher on that trait, they incur an 'affective cost' reflected in increased stress symptoms and less adaptive coping.

Conscientiousness

Consistent with expectation, higher Conscientiousness scores were associated with greater perceived workload, specifically temporal demand. However, contrary to expectation, this relation was observed for one of the easier tasks. A possible reason may be a unique characteristic of the FT task. Recall that in this task dummy rounds were placed in the magazines of officers to induce a weapons jam that required the officers to quickly clear the weapon and then continue firing the remaining rounds. This may have induced time pressure in individuals who were concerned with doing well, i.e., those high in Conscientiousness. Note, however, that such effects were not manifest in the novel tasks, BT and MT, both of which included time pressure as an element. In the case of the MT, the immediate feedback associated with that task may have ameliorated the impact of time pressure on these individuals because the feedback allowed them to track their performance on a shot-by-shot basis. The reason for the absence of any significant association between Conscientiousness and perceived workload for the BT is unclear, and in future research this can be explored by manipulating the different demands within the task.

In addition to workload effects, the expected negative association between Conscientiousness and stress state was observed. Individuals low in Conscientiousness reported greater pre-task distress, results consistent with those of Matthews et al. (1999). However, Matthews and his colleagues also reported a positive association between Conscientiousness and task-engagement. In this study, Conscientiousness was unrelated to pre-task engagement. This may be due to the specific context in which officers completed the tasks in this study. The training tasks were required for the officers by state law, and therefore even those low in Conscientiousness had to have some commitment to task effort.

Agreeableness

The findings for Agreeableness in regard to stress and coping (recall that no prediction was made regarding workload and no association was observed) were generally consistent with expectation. Thus, higher Agreeableness scores were associated with higher post-task engagement and greater task-focused coping. The capacity of individuals higher in Agreeableness for 'going with the flow' and their generally high degree of co-cooperativeness and positive affect may have helped these individuals maintain high task engagement in the face of the demanding training regimen.

Intellect

As in the case of Emotional Stability, the prediction of workload by Intellect was in the direction opposite to that predicted. Higher Intellect scores predicted *greater* perceived workload for the WT, FT, and MT tasks. However, among the five traits, Intellect emerged as the strongest predictor of workload. The characteristics of high Intellect (creativity, adaptability, preference for challenging activities) may influence how these individuals appraised demands in both simpler (FT) and more difficult (MT) shooting tasks, as well as how they appraise their own responses, in terms of effort and perception of own-performance (MT and WT). What is not clear is whether their tendency to seek challenging/ interesting activities and their tendency toward flexibility makes them more sensitive to detecting task demand or simply more willing to acknowledge it. In addition, it is puzzling that Intellect predicted workload for the simplest (e.g., WT) and one of the more challenging (i.e., MT) tasks, but it did not predict the workload associated with the BT task. It may be that of the two more challenging tasks, the BT was sufficiently demanding for all officers across the levels of Intellect, but that the feedback inherent in the MT task induced differential perceptions of workload as a function of Intellect.

Whether these associations generalize beyond police officers to shooting tasks is a matter for future research. However, the results of this study indicate that Intellect, which has previously been found to be weakly related to performance (see Barrick et al., 2001; but see also Burke & Witt, 2002), should not be ignored as a potentially relevant individual differences variable in regard to performance on real-world tasks.

In regard to stress and coping, the results were generally in line with expectation. Higher Intellect scores predicted lower levels of pre-task distress and higher levels of both pre- and post-task engagement. Further, higher Intellect predicted greater use of task-focused coping and less use of emotion focused and avoidant coping strategies. This pattern of results suggests that being high in Intellect may be adaptive for officers under stress, an extension of previous findings in other domains (Penley & Tomaka, 2002). It may be that the greater workload reported by individuals higher in Intellect were not aversive to these individuals (i.e., did not negatively impact stress state) because they tend to seek challenging activities and adapt well to change.

Trait Interactions

Although no significant regressions of performance on traits were observed, a significant Intellect by Conscientiousness interaction was observed for accuracy on the MT task. Performance was superior for individuals who were high on both the Intellect and Conscientiousness scales. On the other hand, those low in Intellect and high on Conscientiousness performed more poorly, suggesting that the threshold for failure of behavioral adaptation under stress (see Hancock & Warm, 1989) may be lower for these individuals relative to those high on both traits. The correlation of Conscientiousness with performance is generally thought to be due to motivational variables (Barrick, Mount, & Judge, 2001). However, it may be

that such associations can only emerge if the individuals also have the cognitive flexibility to adapt to environmental demands. However, the current results suggest that these effects are also task dependent. This interaction was observed only for the MT task, possibly because that task was particularly unfamiliar (e.g., no standard paper targets) for these officers. By contrast, the other three tasks were relatively more familiar. Although the BT task was relatively novel in that it differed from standard shooting range tasks (such as the WT), it involved shooting at paper silhouette targets familiar to the officers.

For time perception, significant regressions were observed for the interaction of Intellect with Emotional Stability and of Emotional Stability with Extraversion. In each case the interaction was found for the BT task. The interaction of Intellect and Emotional Stability was consistent with expectation, in that individuals low on both traits made larger time estimation errors relative to those low in Intellect but high in Emotional Stability (see Figure 4). For the Emotional Stability by Extraversion interaction, differences in Extraversion were larger for those lower in Emotional Stability, as expected (see Figure 1). However, the direction of these associations were opposite to that expected. Recall that it was hypothesized that for those low in Emotional Stability, those also low in Extraversion would produce greater errors than those high in Extraversion. The reverse was observed: Individuals low in Extraversion produced more accurate time estimates than those high on that trait. This interaction is somewhat puzzling, given that the lower basal arousal associated with higher Extraversion should lend itself to improving performance (i.e., more accurate time estimates) on a task requiring short term memory capacity (see Humphreys & Revelle, 1984). It may result from greater attention among those high in Extraversion to external cues, making time appear to pass more slowly (Hancock & Weaver, 2004). High Emotional Stability may attenuate these associations, while low Emotional Stability

may enhance them because of higher levels of limbic activation (Eysenck, 1967; Eysenck & Eysenck, 1985). Another possibility is that the overestimation by individuals low in Emotional Stability and high in Extraversion is a manifestation of impulsivity, which has been shown to be associated with overestimation of time intervals (Lawrence & Stanford, 1999; but see also Lennings & Burns, 1998). Future research can explore this possibility by measurement of the facets of the five factors, and analysis of interactions among the facets, particularly interactions involving impulsivity.

To the degree that distortions of time perception represent the narrowing of attention (Hancock & Weaver, 2004), these results suggest that those low in Emotional Stability do indeed have less capacity to devote to temporal stimuli, possibly due to the resources required to cope with their affective state. However, this association only emerges if the individual is also high in Extraversion or low on Intellect. In the case of the latter, the lower degree of adaptability and creativity associated with low Intellect may have constrained the number of alternative strategies for estimating time intervals (Rammsayer, 2002), thereby exacerbating the influence of Emotional Stability.

Interactions were also observed for regressions of post-task distress and pre-task worry. Results for the Intellect by Conscientiousness interaction associated with post-task distress were consistent with the expectation that differences in distress across levels of Conscientiousness were larger among those lower in Intellect. However, contrary to prediction, among the individuals low in Intellect, it was those *higher* in Conscientiousness that experienced greater distress relative to individuals low on that trait. While this is opposite the trend reported by Matthews et al. (1999), it is important to note that they did not test for trait interactions in their study. The combination of low Intellect and high Conscientiousness might make such individuals

vulnerable to distress because of the demands placed on officers during the session. Recall that distress reflects appraisals of capacity overload (Matthews et al., 2002). Substantial task demands in a professional training environment would more likely facilitate appraisals of stress among individuals higher in Conscientiousness given their tendency for achievement striving and task focus. That is, these individuals may have had less capacity to cope with stress (low Intellect) but simultaneously cared more about their performance (high Conscientiousness) and therefore experienced higher levels of distress.

The regression of post-task distress on the interaction of Emotional Stability and Extraversion was also in line with the expectation that differences in Extraversion might emerge for those low in Emotional Stability. However, contrary to prediction, among individuals low in Emotional Stability it was those *high* in Extraversion who experienced the higher levels of distress. It is interesting that these same individuals also overestimated duration in the BT task (see Figure 1), suggesting a link between time distortion and distress. However, this link must remain very tentative because the findings for time estimation were task specific, and in this study we were unable to assess stress state after each task. A more fine-grained measurement of stress state (i.e., after each task) would be advisable in order to determine whether there is a task-specific link between errors in time estimation and distress for individuals low in Emotional Stability and high in Extraversion.

In regard to pre-task worry, an interaction between Intellect and Emotional Stability was observed. Thus, among individuals low in Intellect, those who were also low in Emotional Stability reported greater worry than those higher in Emotional Stability (see Figure 6). It is interesting that this interaction was restricted to pre-state worry. This indicates that the pattern of results are stronger for the *anticipation* of task performance rather than for the performance

itself, since similar differences observed in post-task worry were not significant after the pre-state was accounted for in the model. Note that worry reflects primarily the cognitive component of anxiety rather than the affective component (Matthews et al., 1999; 2002). Thus, the link between stress and the interaction of Intellect and Emotional Stability is manifested cognitively rather than affectively.

Patterns Across Traits

Performance. In general, the results of the present work indicate that traits predict performance on specific tasks, but only in the context of interactions with other traits. Barrick et al. (2001) have noted that the mechanisms by which traits influence job performance are not well articulated. For specific task performance, mechanisms have been proposed for Extraversion and Emotional Stability (Eysenck, 1967; Eysenck & Eysenck, 1985), but evidence has not unambiguously supported these as general mechanisms (e.g., see Matthews et al., 2000; 2003). Thus, given that few studies have investigated these interactions in the context of human performance; future research should explore the mechanisms that govern the interactions between traits and their relation to task performance. A fruitful approach would be to explore the relations between task characteristics (e.g., challenge, difficulty, and complexity) and the link between performance and traits.

Workload. In regard to workload, the finding that temporal demand emerged as the dominant workload scale in regard to the trait regressions suggests that appraisals of the temporal components of task are more closely linked to personality differences than other workload dimensions. This may be due to the trait differences associated with the perception of time (Rammsayer, 2002; Zakay, Lombranz, and Kziniz, 1984). The data also show, however, that these relations depend on the nature of the task. Given the relatively small database available,

systematic research relating traits and task characteristics to perceived workload is warranted.

Stress State. In regard to stress state, the pattern of associations with the big five factors is generally consistent with those reported in previous research (Matthews et al., 1999). Thus, previous findings using student and office worker populations extend to police officers. However, the differences in stress-state as a function of traits are not uniform across tasks and traits. In addition, the association between traits and stress varied according to stress scale, confirming the multidimensionality of the stress concept itself (Matthews, 2001). Pre-task distress emerged as the scale with the strongest association with trait, with only Agreeableness failing to predict this component of pre-task state. By contrast, worry was the dimension least associated with traits, with a statistically significant regression only emerging in the context of an interaction between Intellect and Emotional Stability. Thus, for the participants in this study, traits were more strongly related to the affective and motivational/energetic dimensions of stress state rather than to the cognitive component of stress. This differs somewhat from the findings of Matthews et al. (1999), underscoring that one cannot assume that the relation of traits to states will be uniform across populations or types of task. Future research to identify how the pattern of trait-state links varies according to task demands is a logical next step and could contribute to the improvement of personality theory (Matthews et al., 2003) and improved understanding of individual differences in performance.

Note that a limitation of the current study is the absence of control over the task environment. Measurement of post-task stress state at the end of the session (after completing all four tasks) may have attenuated differences in post-task state as a function of traits due to the demanding nature of the set of tasks. Future research, both in the field and in simulation, can circumvent these difficulties by use of a shorter version of the DSSQ now available (Helton,

2004). Future studies can examine the stress state associated with individual task performance. This could be potentially important given that the relation between trait and workload varied as a function of task in the current study. It is likely that the same will be true of stress state.

Coping. The findings in regard to coping indicate that high levels of Emotional Stability, Agreeableness, and Intellect may be adaptive for officers' under stress, an extension of previous findings (e.g., Penley & Tomaka, 2002). The absence of adaptive coping for Conscientiousness was not expected, and of the three coping strategies avoidant coping was the least strongly related to Agreeableness, Conscientiousness, or Emotional Stability. This pattern differs from other studies, and suggests that how traits influence coping in general or in the social environments may not extend to the training tasks of the sort experienced by police officers. The pattern of results also varies somewhat from findings related to coping with general life stress (e.g., McCrae & Costa, 1986) indicating that general stress coping strategies may not be manifested in specific task environments. This suggests the need for coping assessment that addresses task coping specifically rather than general strategies individuals use to cope with stress (see Matthews & Campbell, 1998).

Summary and Conclusions

The results of this study indicate that the relation of traits to performance, workload, stress, and coping are complex and depend on both task characteristics and, in the case of workload and stress, the dimension measured. This should come as no surprise, given that the criterion variables, particularly workload and stress, are themselves multidimensional. Indeed, these results also indicate the importance of measuring perceived workload and stress state in addition to performance, because they provide additional vistas into the person's status beyond that afforded by performance measurement alone. Further, in some instances traits were

associated with performance or stress (e.g., the worry dimension) only in the context of an interaction with another trait. These results underscore the importance of considering interactions between traits when exploring individual differences in performance (see Burke & Witt, 2002). The influence of a particular trait on performance and stress outcomes may, in some circumstances (e.g., complex, real world task performance), depend on the 'context' of other traits. Further, identification of how trait profiles relate to performance outcomes as well as stress and perceived workload may prove useful for assessment of these variables in real-world contexts. Hence, future research should incorporate tests for trait interactions to avoid missing potential trait-performance links that may be masked when traits are considered in isolation. Note that these interactions, when they were observed, also varied according to task or, in the case of stress, the dimension measured. It is therefore crucial that future research identify the task characteristics that determine when these associations occur, and the mechanisms by which traits and states intact with the task environment to influence behavioral response.

Acknowledgement

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621, P.A. Hancock, *Principal Investigator*. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant. The authors also wish to express their gratitude to the Heads of the Police Departments who let us perform the research with their officers, to the 71 police officers who volunteered to participate in the study, and to Adams Greenwood-Ericksen and Jennifer M. Ross, who helped in data collection and reduction.

References

Adlam, K.R. (1982). The police personality: Psychological consequences of being police officers. *Journal of Police Science and Administration, 10*, 344-349.

Alexander, D.A., Inees, G., Irving, B.L., Sinclair, S.D., & Walker, L.D. (1991). *Health, stress, and policing.* London: The Police Foundation.

Barrick, M.R., & Mount, M.K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Barrick, M.R., Mount, M.J., & Judge, T.A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Personality and Performance, 9*, 9-30.

Beutler, L., Nussbaum, P.D., & Meredith, K.E. (1988). Changing personality patterns of police officers. *Professional Psychology Research and Practice, 88*, 503-507.

Biggam, F.H., & Power, K.G. (1996). The personality of the Scottish police officer: The issue of positive and negative affectivity. *Personality and Individual Differences, 20,* 661-667.

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review, 4*, 184-197.

Block, R.A., Zakay, D., & Hancock, P.A. (1999). Developmental changes in human duration judgments: A meta-analytic review. *Developmental Review, 19,* 183-211.

Brown, J.M., & Campbell, E.A. (1990). Sources of occupational stress in the police. *Work and Stress, 4*, 305-318.

Brown, J.M., & Campbell, E.A. (1994). *Stress and policing: Sources and strategies.* Chichester, UK: Wiley.

Burbeck, E., & Furnham, A. (1985). Police officer selection: A critical review of the literature. *Journal of Police Science and Administration, 13*, 58-69.

Burke, L.A., & Witt, L.A. (2002). Moderators of the openness to experience-performance relationship. *Journal of Managerial Psychology, 17*, 712-721.

Byers, J.C., Bittner, A.C., & Hill, S.G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In: A. Mital (Ed.), *Advances in industrial ergonomics and safety, I* (pp. 481-485). London: Taylor & Francis.

Carpenter, B.N., & Raza, S.M. (1987). Personality characteristics of police applicants: Comparisons across subgroups and with other populations. *Journal of Police Science and Administration, 15*, 10-17.

Damos, D.L. (1988). Individual differences in subjective estimates of workload. In: P.A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 231-237). Amsterdam: Elsevier.

Damos, D.L., & Bloem, K. (1985). Type A behavior pattern, multiple-task performance, and subjective estimation of mental workload. *Bulletin of the Psychonomic Society, 23*, 53-56.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*, 183-201.

Endler, N.S., & Parker, J.D.A. (1990). Multidimensional assessment of coping: A critical evaluation. *Journal of Personality and Social Psychology, 58*, 844-854.

Eysenck, H.J. (1967). *The biological basis of personality*. Springfield, IL: Charles C. Thomas.

Eysenck, M.W., & Calvo, M.G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*, 409-434.

Eysenck, H.J., & Eysenck, M.W. (1985). *Personality and individual differences: A natural*

science approach. New York: Plenum Press.

Fairclough, S.H. (1991). Adapting the TLX to measure driver mental workload (Report V1017/BERTIE/No. 71). Loughborough, Leics, UK: HUSAT Research Institute.

Goldberg, L.R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4*, 26-42.

Grubb, P.L., Miller, L.C., Nelson, W.T., Warm, J.S., Dember, W.N., & Davies, D.R. (1994). Cognitive failure and perceived workload in vigilance performance. In: M. Mouloua and R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 115-121). Hillsdale, NJ: Erlbaum.

Gudjonsson, G.H., & Adlam, R. (1983). Personality patterns of British police officers. *Personality and Individual Differences, 5*, 507-512.

Hancock, P.A. (1996). Effects of control order, augmented feedback, input device and practice on tracking performance and perceived workload. *Ergonomics, 39*, 1146-1162.

Hancock, P.A., & Dirkin, G.R. (1983). Stressor induced attentional narrowing: Implications for design and operation of person-machine systems. *Proceedings of the Human Factors Association of Canada, 16*, 19-21.

Hancock, P.A., & Hart, S.G. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in Design, 10,* 6-16.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519-537.

Hancock, P.A., & Weaver, J.L. (2004). On time distortion under stress. *Theoretical Issues in Ergonomics Science, 6*, 193-211.

Hart, S. G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results

of empirical and theoretical research. In: P.A. Hancock and N. Meshkati (Eds.), Human mental workload (pp. 139-183). Amsterdam: Elsevier.

Helton, W.S. (2004). Validation of a short stress state questionnaire. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*, 1238-1242.

Humphreys, M.S., & Revelle, W. (1984). Personality, motivation, and performance: A theory of the relationship between individual differences and information processing. *Psychological Review, 91*, 153-184.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression, 2nd ed*. Thousand Oaks: Sage.

Kardum, I., & Hudek-Knezevic, J. (1996). The relationship between Eysenck's personality traits, coping styles, and moods. *Personality and Individual Differences, 20*, 341-350.

Kirkcaldy, B.D. (1984). Individual differences in time estimation. *International Journal of Sports Psychology, 15*, 11-24.

Lawrence, J.B., & Stanford, M.S. (1999). Impulsivity and time of day: Effects on performance and cognitive tempo. *Personality and Individual Differences, 26*, 199-207.

Lazarus, R.S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer-Verlag.

Lennings, C.J., & Burns, A.M. (1998). Time perspective: Temporal extension, time estimation, and impulsivity. *The Journal of Psychology, 132*, 367-380.

McCrae, R.R., & Costa, P.T. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individuals Differences, 6,* 587-597.

McCrae, R.R., & Costa, P.T. (1986). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81-90.

McCrae, R.R., & Costa, P.T. (1999). A five-factor theory of personality. In: L.A. Pervin and O.P. John (Eds.), *Handbook of personality: Theory and research, 2nd ed.* (pp. 139-153). New York: Guilford Press.

Matarazzo, J.D., Allen, B.V., Saslow, G., & Wiens, A.N. (1964). Characteristics of successful policemen and firemen applicants. *Journal of Applied Psychology, 48*, 123-133.

Matthews, G. (2001). Levels of transaction: A cognitive science framework for operator stress. In: P.A. Hancock and P.A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 5-33). Mahwah, NJ: Erlbaum.

Matthews, G., & Campbell, S.E. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society, 42*, 821-825.

Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., & Warm, J.S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion, 2*, 315-340.

Matthews, G., Davies, D.R., Westerman, S.J., & Stammers, R.B. (2000). *Human performance: Cognition, stress, and individual differences*. London: Psychology Press.

Matthews, G., Deary, I.J., & Whiteman, M.C. (2003). *Personality traits, 2nd ed*. Cambridge: Cambridge University Press.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state 'big three'? In: I. Mervielde, I.J. Deary, F. De Fruyt, and F. Ostendorf (Eds.), *Personality Psychology in Europe* (vol 7. pp. 335-350). Tilburg, the Netherlands: Tilburg University Press.

Nygren, T.E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors,*

33, 17-33.

Pedhazur, E.J. (1997). *Multiple regression in behavioral research, 3rd ed*. Ft. Worth: Harcourt Brace.

Penley, J.A., & Tomaka, J. (2002). Associations between the big five, emotional responses, and coping with acute stress. *Personality and Individual Differences, 32,* 1215-1228.

Rammsayer, T.H. (2002). Temporal information processing and basic dimensions of personality: Differential effects of psychoticism. *Personality and Individual Differences, 32*, 827-838.

Robertson, M. (1984). Personality differences as a moderator of mental workload behavior: Mental workload performance and strain reactions as a function of cognitive complexity. *Proceedings of the Human Factors Society Annual Meeting, 28*, 690-694.

Robertson, M., & Meshkati, N. (1985). Analysis of the effects of two individual differences classification models on experiencing mental workload of a computer generated task: A new perceptive to job design and task analysis. *Proceedings of the Human Factors Society Annual Meeting, 29,* 178-182.

Rose, C.L., Murphy, L.B., Byard, L., & Nikzad, K. (2002). The role of the big five personality factors in vigilance performance and workload. *European Journal of Personality, 16*, 185-200.

Salgado, J.F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82*, 30-43.

Thompson, B., Kirk-Brown, A., & Brown, D. (2001). Women police: The impact of stress on family members. In: P.A. Hancock and P.A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 200-210). Mahwah, NJ: Erlbaum.

Zakay, D. (1998). Attention allocation policy influences prospective timing. *Psychonomic*

Bulletin & Review, *5*, 114-118.

Zakay, D., & Block, R.A., (2004). Prospective and retrospective duration judgments: An executive-control perspective, *Acta Neurobiolgiae Experimentalis*, *64*, 319-328

Zakay, D., Lomranz, J., & Kazinz, M. (1984). Extraversion-introversion and time perception. *Personality and Individual Differences, 5*, 237-239.

Table 1

Predictions for regressions on the five traits.

Criterion	E	ES	C	A	I
Shooting Accuracy	+	+	+	+	
Time Estimation	+	+	+	+	+
Workload	-	-	+		-
Stress	-	-	-	-	-
Coping	+	+	+	+	+

Note. + higher level on trait will predict higher scores for performance, workload, and stress, and predict a pattern of more effective coping; - higher level on trait will predict lower scores for performance, workload, and stress, and predict a pattern of less effective coping; no entry= no prediction made. E=Extraversion; ES=Emotional Stability; C=Conscientiousness; A=Agreeableness; I=Intellect.

Table 2

Descriptive statistics for the trait and DSSQ state measures.

Traits	Mean	SD	N
Extraversion	119.05	19.12	60
Emotional Stability	112.00	15.42	59
Conscientiousness	139.45	17.22	58
Agreeableness	132.62	19.32	61
Intellect	123.16	20.88	61
State (Pre)			
Task Engagement	1.22	.71	65
Distress	-1.24	.83	65
Worry	-.12	.88	65
State (post)			
Task Engagement	1.67	.86	65
Distress	-.53	1.28	65
Worry	-.14	.96	65

Table 3

Descriptive statistics for performance and time estimation measures.

Task	N	Shooting Accuracy	Time Estimation (DJR)	Time Estimation (AE)
WT	71	74.73 (18.75)	----	----
FT	71	65.61 (21.63)	----	----
BT	71	58.31 (20.34)	1.08 (.68)	.58 (.74)
MT	71	62.98 (15.24)	1.16 (.70)*	.57 (.71)*

Note. Standard deviations are in parentheses. WT=Warm-up task; FT=Flashlight task;

BT=Barrel task; MT= Metal task; DJR=Duration judgment ratio; AE= Absolute error.

* For these measures N=69.

Table 4

Descriptive statistics for RTLX workload scales.

Task	N	GWL	MD	PD	TD	P	Eff	F
WT	67	42.53 (15.10)	51.15 (26.03)	28.22 (21.13)	36.96 (25.95)	63.06 (23.95)	47.46 (23.67)	28.31 (22.89)
FT	68	48.40 (16.76)	57.34 (24.43)	39.51 (26.45)	43.88 (26.99)	61.53 (23.86)	52.00 (25.12)	36.12 (23.91)
BT	68	59.65 (16.01)*	67.96 (20.87)	61.89 (23.40)	65.30 (22.88)	55.89 (22.84)	63.90 (21.29)	44.46 (28.49)*
MT	70	56.51 (17.84)**	64.21 (24.91)	52.86 (25.29)	61.29 (25.36)	62.98 (15.24)***	60.84 (21.78)	47.22 (26.24)**

Note. Standard deviations are in parentheses. WT=Warm-up task; FT=Flashlight task; BT=Barrel task; MT= Metal task; GWL=Global Workload; MD=Mental Demand; PD=Physical Demand; TD=Temporal Demand; P=Performance; Eff=Effort; F=Frustration; E=Extraversion; A=Agreeableness; C=Conscientiousness; ES=Emotional Stability; I=Intellect.

*for this scale N=68; ** for this scale N=69; *** for this scale N=71.

Table 5

Outcomes of regressions on the five traits.

Criterion	E	ES	C	A	I
Shooting Accuracy	0	0	0	0	0
Time Estimation	0	0	0	0	0
Workload	0	+#	+*	0	+#
Stress	-*	-*	-*	-*	-*
Coping	0	+*	0	+*	+*

Note. + higher level on trait predicted higher scores for performance, workload, and stress, or predicted a pattern of more effective coping; - higher level on trait predicted lower scores for performance, workload, and stress, or predicted a pattern of less effective coping; 0 = no effect observed. * Outcome generally consistent with prediction; # Outcome opposite of that predicted. E=Extraversion; ES=Emotional Stability; C=Conscientiousness; A=Agreeableness; I=Intellect.

Figure Captions

Figure 1. Duration judgment ratio for the BT task as a function of emotional stability for three levels of extraversion.

Figure 2. Post-task distress as a function of emotional stability for three levels of extraversion.

Figure 3. Accuracy on the MT task as a function of intellect at three levels of conscientiousness.

Figure 4. Duration judgment ratio for the BT task as a function of intellect for three levels of emotional stability.

Figure 5. Post-task distress as a function of intellect at three levels of conscientiousness.

Figure 6. Pre-task Worry as function of Intellect at three levels of Emotional Stability.

Figure 1

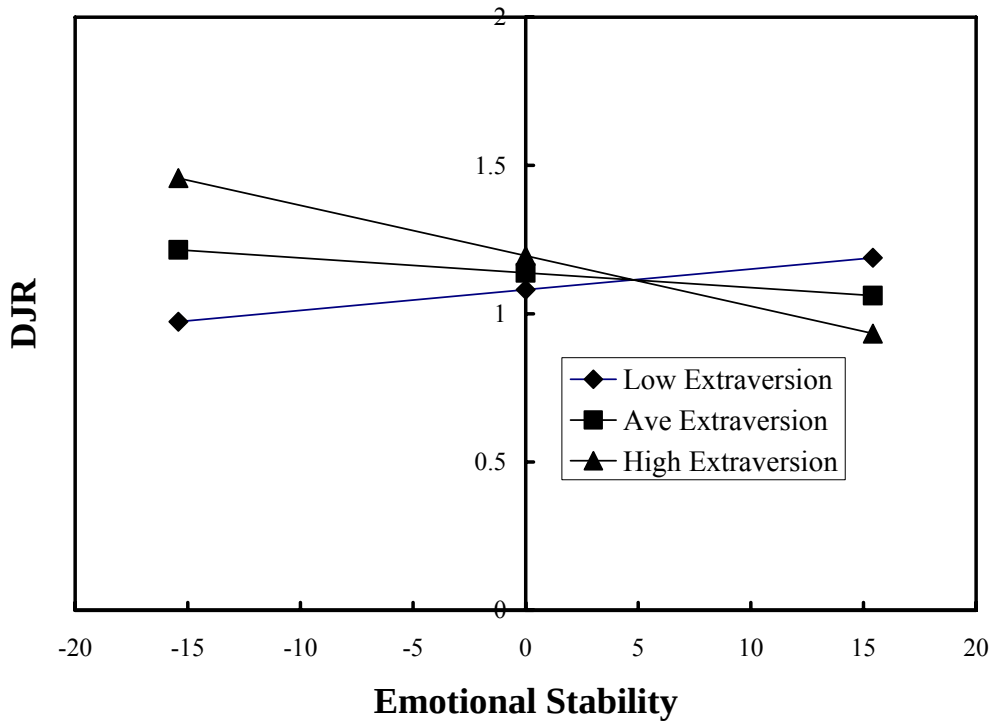

DJR
2
1.5
1
0.5
0
-20
-15
-10
-5
0
5
10
15
20
Emotional Stability
Low Extraversion
Ave Extraversion
High Extraversion

Figure 2

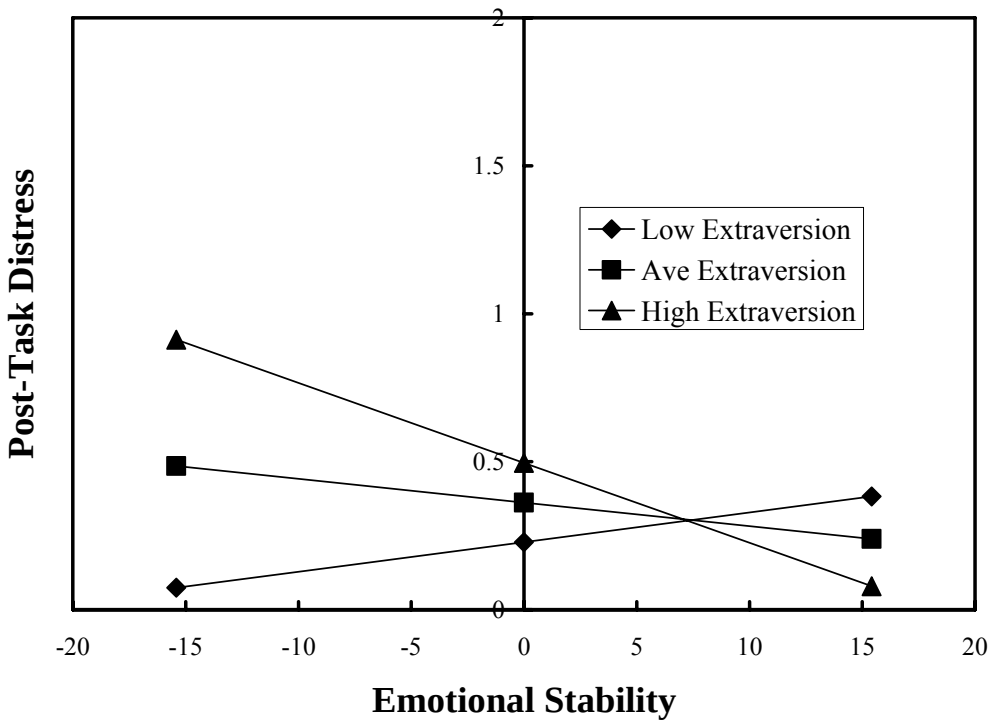

Post-Task Distress
2
1.5
1
0.5
0
-20
-15
-10
-5
0
5
10
15
20
Emotional Stability
Low Extraversion
Ave Extraversion
High Extraversion

Figure 3

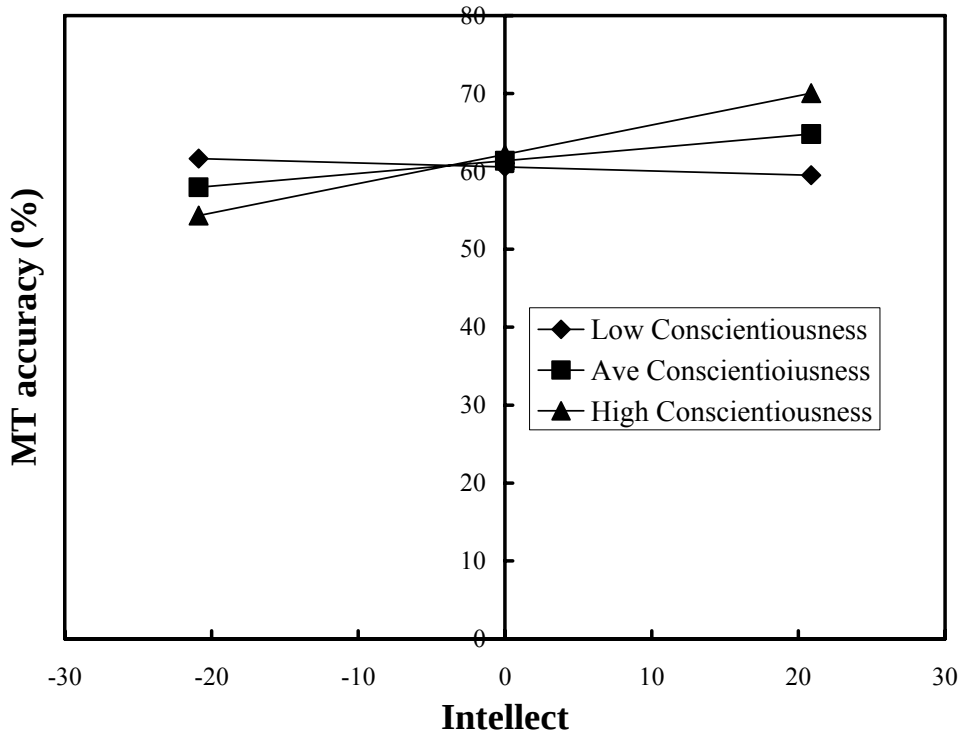

MT accuracy (%)
80
70
60
50
40
30
20
10
0
-30
-20
-10
0
10
20
30
Intellect
Low Conscientiousness
Ave Conscientioiusness
High Conscientiousness

Figure 4

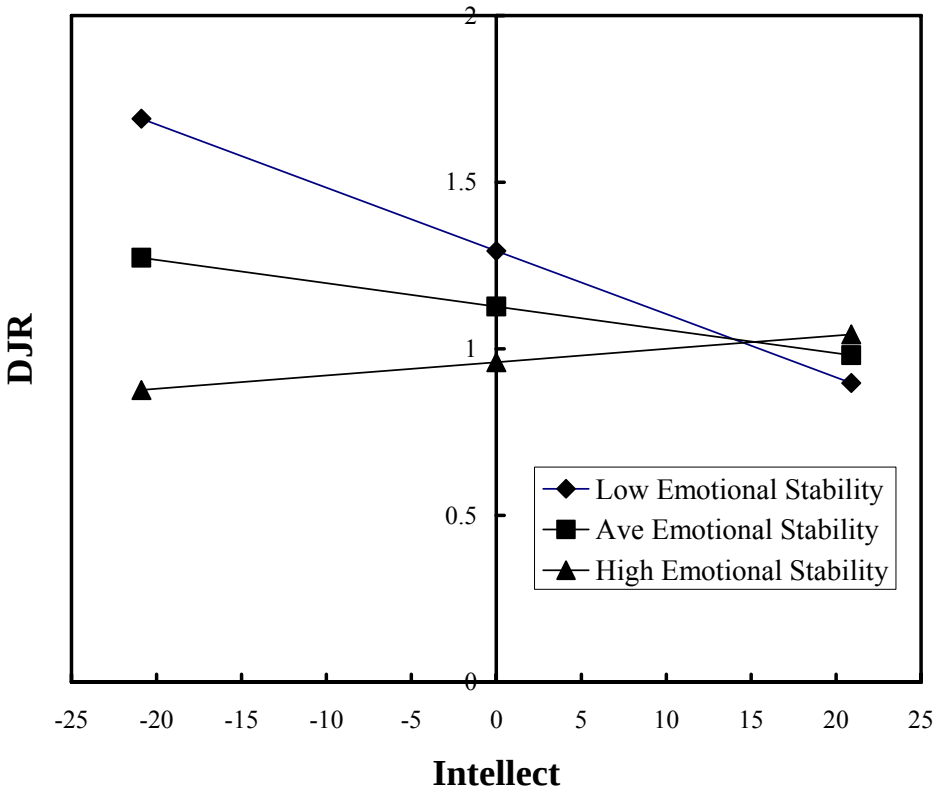

DJR
2
1.5
1
0.5
0
-25
-20
-15
-10
-5
0
5
10
15
20
25
Intellect
Low Emotional Stability
Ave Emotional Stability
High Emotional Stability

Figure 5

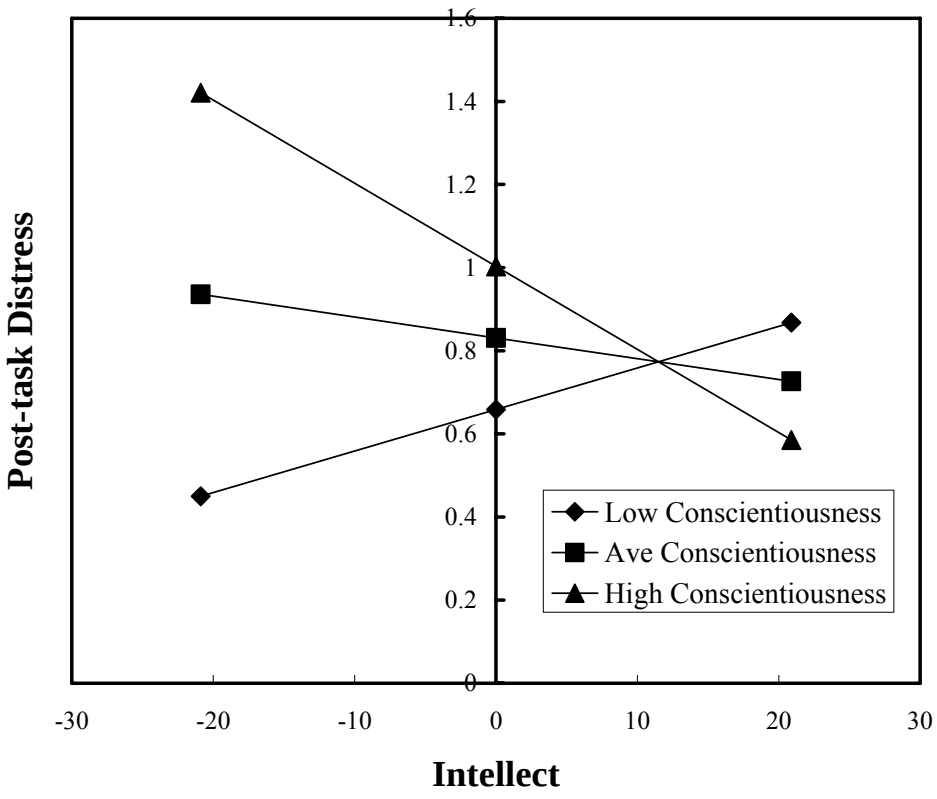
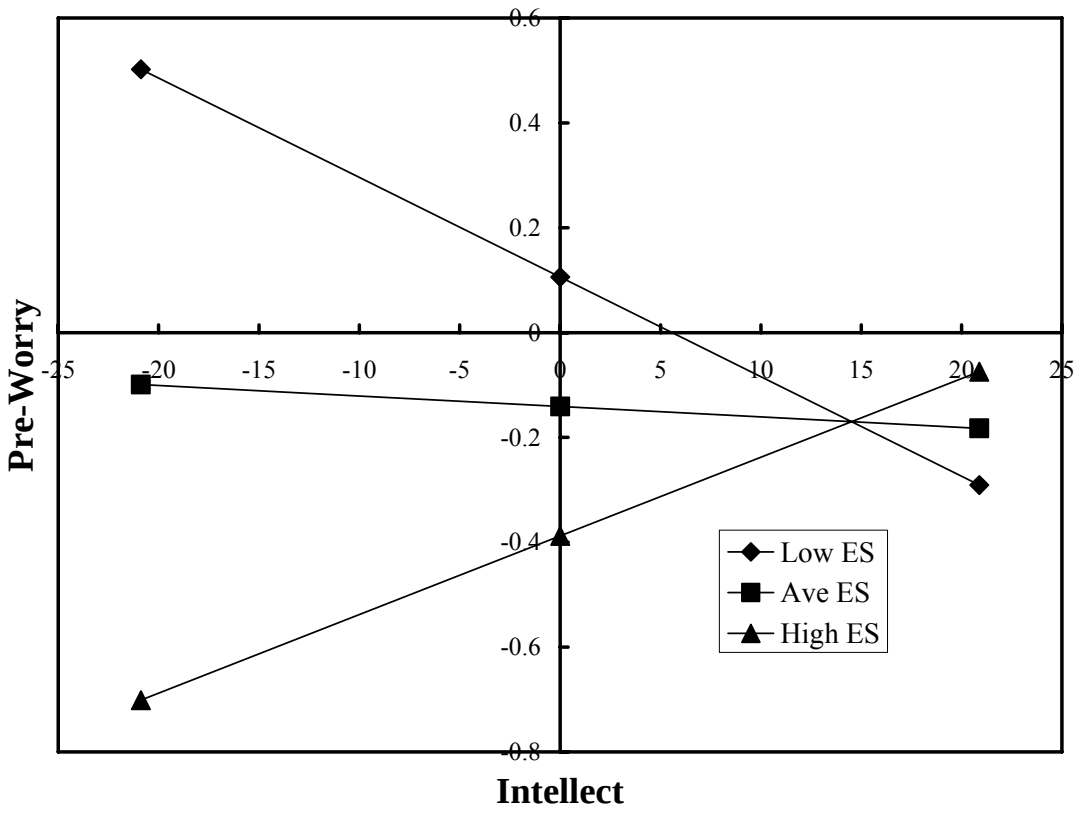

Post-task Distress
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-30
-20
-10
0
10
20
30
Intellect
Low Conscientiousness
Ave Conscientiousness
High Conscientiousness

Figure 6


Pre-Worry
Intellect
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-25
-20
-15
-10
-5
0
5
10
15
20
25
Low ES
Ave ES
High ES

Training for Vigilance: Using Predictive Power to Evaluate Feedback Effectiveness

James L. Szalma and Peter A. Hancock

University of Central Florida, Orlando FL

Joel S. Warm, William N. Dember, and Kelley S. Parsons

University of Cincinnati, Cincinnati, OH

Address correspondence to James L. Szalma, Institute for Simulation and Training, University of Central Florida, PO Box 161390, Orlando, FL 32816-1390; jszalma@mail.ucf.edu

Running Head: Knowledge of Results in Vigilance

Keywords: Vigilance; Sustained Attention; Training; Knowledge of Results; Performance

ABSTRACT

Objective

We examined the effects of Knowledge of Results (KR) on vigilance accuracy, and report the first use of positive and negative predictive power (PPP/NPP) to assess vigilance training effectiveness.

Background

Training individuals to detect infrequent signals amongst a plethora of non-signals is critical to success in many failure-intolerant monitoring technologies. KR has been widely used for vigilance training, but the effect of KR-schedule on accuracy has been neglected. Previous research on training for vigilance has used signal detection metrics or hits and false alarms. In this study diagnosticity measures were applied to augment traditional analytic methods.

Method

We examined the effects of continuous knowledge of results (KR) and a partial-KR regimen versus a no-KR control on decision diagnosticity.

Results

SDT analysis indicated that KR induced conservatism in responding but did not enhance sensitivity. However, KR in both forms equally enhanced PPP while selectively impairing NPP.

Conclusion

There is a tradeoff in the effectiveness of KR in reducing false alarms and misses. Together, SDT and PPP/NPP measures provide a more complete portrait of performance effects.

Application

PPP and NPP together provide another assessment technique for vigilance performance, and as an additional diagnostic tool these measures are potentially useful to the human factors community.

INTRODUCTION

Vigilance refers to the ability of observers to detect infrequent signals over prolonged periods on watch (Davies & Parasuraman, 1982). The importance of vigilance has vaulted to the forefront of current social concerns regarding detection of terrorist activities (Hancock & Hart, 2002). For example, the screening devices deployed by the Transportation Security Administration (TSA) to screen 100% of checked baggage represents a major vigilance requirement for operators (Harris, 2002). Sustained attention is also central to the majority of advanced human-machine systems, including; air-traffic control, cockpit monitoring, industrial quality control, nuclear power generation, medical monitoring, cytological screening (Warm, 1993). One perennial concern remains how to train individuals for vigilance and how to evaluate performance. An effective training procedure is to provide information about an operator's responses via knowledge of results (KR). Research has indicated that monitors trained with KR perform better after the training aid has been withdrawn than controls who received no KR training (see Davies & Parasuraman, 1982; Warm & Jerison, 1984). However, the mechanisms by which KR exerts its effects, and how best to employ KR for training, remains unspecified.

In vigilance, performance accuracy is assessed predominantly by evaluating the proportion of correct detections and false alarms. These scores are used to compute signal detection theory (SDT) measures of sensitivity (e.g., d′ or A′; Macmillan & Creelman, 2005; See, Howe, Warm, & Dember, 1995) and response bias (e.g., β; Davies & Parasuraman, 1982; and see See, Warm, Dember, & Howe, 1997). Sensitivity assesses operators' ability to discriminate signal from non-signal. In vigilance, response bias, which tends to become increasingly conservative over time, is thought to reflect increasing awareness on behalf of operators that the signals for which they are searching are relatively rare (Craig, 1978). Indeed, KR has been shown to enhance perceptual sensitivity while increasing β (e.g., see Szalma, Miller, Hitchcock, Warm, & Dember, 1998). However, while these established measures assess critical aspects of performance, they do not directly capture the diagnostic accuracy of response, i.e., the proportion of 'yes' or 'no' responses by individuals that are correct. SDT measures reflect diagnosticity indirectly, since one might infer greater diagnostic accuracy from higher d' scores. It is possible, however, to achieve high sensitivity while making poorer diagnostic decisions

(Parasuraman, Hancock, & Olofinboba, 1997). This is particularly evident when the *a-priori* probability of a signal (base rate) is very low, as in most real-world situations. Indeed, Parasuraman, Hancock, and Olofinboba (1997) demonstrated that under these conditions the accuracy of system response can also be low even when sensitivity is exceptionally high. An operator with a high sensitivity can generate unacceptably large numbers of false alarms and reduce their diagnostic power, which inevitably results in reduced overall system effectiveness.

It is therefore critical in operational settings that observers learn to be highly diagnostic as well as sensitive to perceptual changes in the displays which they are monitoring. Specifically, it is vital that when an operator indicates a signal is present then the signal actually is present. Conversely, when an operator indicates that a signal is not present, it is vital that the signal is really not there. These aspects of performance are captured by the decision theory measures of positive predictive power (PPP), which is the proportion of 'yes' responses that are actually correct, and negative predictive power (NPP), which is the proportion of 'no' responses that are actually correct. The computational formula for PPP is H/(H+FA), where H= number of signals detected (hits) and FA=number of false alarms. The comparable computational formula for NPP is CR/(CR+M), where CR=number of correct rejections and M=number of signals missed. These indices are frequently employed in evaluating decision-making in medicine (e.g., see Linton, 1996), but have rarely been employed in the human factors literature (for an exception, see Getty, Swets, Pickett, & Gonthier, 1995). A perfectly accurate observer would yield a PPP of 1.0; a score of zero would indicate no correct detections and no diagnosticity. Similarly, an observer who correctly rejected all non-signals and committed no misses would achieve a NPP score of 1.0, while a score of zero indicates that no correct rejections were made.

Given the importance of training personnel for monitoring and the relative costs of false alarms and misses, it is vital that any training regimen maximize both sensitivity and diagnostic power. Hence, we applied PPP and NPP to the examination of the effectiveness of KR for vigilance training, and we compared these measures to non-parametric measures of SDT. We do not propose PPP/NPP measures replace SDT, which is one of the most powerful and useful quantitative theories in behavioral science. Rather, we show

how inclusion of these 'new' measures complements SDT indices and provides an alternative vista into detection capability.

Knowledge of results has been shown to be an effective training aid for a variety of tasks beyond vigilance, including those entailing motor skills (e.g., Salmoni, Schmidt, & Walter, 1984). However, the schedule of KR presentation influences the effectiveness of such training. Specifically, partial-KR, i.e., KR provided only during portions of training, has been shown to be more effective than continuous-KR in training motor skills (Salmoni, Schmidt, & Walter, 1984). If these findings extend to vigilance, performance during training should be superior for monitors receiving continuous-KR relative to those who receive partial-KR, but the pattern should reverse during the transfer phase. Such a result would suggest that partial-KR would be the more effective training mode for vigilance in operational settings.

Most vigilance studies employing KR for training have used a continuous-KR schedule, but a limited number of studies have employed partial-KR (i.e., McCormack, Binding, & McElheran, 1963; Warm, Hagner, & Meyer, 1971). In the latter work, Warm, Hagner, and Meyer (1971) found that observers provided with KR only 50% of the time during training showed a smaller vigilance decrement in a subsequent test period without KR than observers who had received continuous-KR during training. Warm et al. attributed this partial-KR advantage to the motivational effects of feedback. However, they employed a speed measure, *reaction time to correct detections*, to assess performance in a simple vigilance task -- the onset of a small jewel light, and thus no study has examined the effect of partial-KR on vigilance in which the *accuracy* of operator responses was the critical measure of interest. While speed is of importance in detection tasks, it is arguable that accuracy is a more important real-world imperative. Indeed, prior research has shown that accuracy and response time may not always reflect the same perceptual processes (Santee & Egeth, 1982). Thus, effects of partial-KR on response accuracy in vigilance have yet to be evaluated and this is another goal, unique to the present study.

METHOD

Sixteen observers (8 men and 8 women) were assigned at random to each of three KR groups

(continuous-KR, partial-KR, and a no-KR control). Observers ranged in age from 18 to 40 (mean= 20.5), and served to fulfill a University course requirement. All observers had normal or corrected-to-normal vision and were reportedly free of hearing impairments. They took part in a 45-min session divided into a training phase of three continuous 4-min periods of watch followed by a test phase of five continuous 4-min periods. A ten-minute rest interval separated training and testing phases.

The display consisted of a 1.4 cm diameter green disk flanked on each side by a 1.0 cm vertical green line appearing against a gray background. The lines were connected to the disk by a 1.0 cm horizontal green line that extended from the midpoint of each vertical line through the horizontal diameter of the disk. Neutral events, requiring no overt response, were cases in which the two vertical lines were equidistant from the disk. Critical signals were cases in which one vertical line was 0.4 cm farther from the disk than the other line. Observers were instructed to respond by pressing the spacebar on a computer keyboard whenever a critical signal appeared on the screen. Responses occurring within 1.3 sec of critical signal onset were recorded by the computer as correct detections. All other responses were recorded as false alarms. The screen was mounted on a table at eye level approximately 54 cm from the seated observer. This task was selected in part because of existing comparable results readily available to the authors (e.g., Szalma et al., 1998).

In both experimental phases, the disk/line display was presented in the center of the screen for 200 msec once every 2 sec (event rate = 30/min). Critical signals occurred twice/min (signal probability = 0.066) at random intervals within each period on watch. Half of the signals in each period were cases in which the left vertical line was farther from the disk, and the other half were cases in which the right vertical line was farther from the disk. The order in which the two types of critical signal were presented (left vs. right) was randomly distributed over each watch period. During training, KR was provided by the computer using a pre-recorded female voice to announce correct detections ("Correct"), missed signals ("Miss"), and errors of commission ("False Alarm"). In the No-KR condition, the voice announced "saved" after each response. In the partial-KR group, participants were informed that KR would be provided by means of the voice which announced “feedback on” at the beginning of periods 1 and 3, and “feedback off” at the beginning of period

2. Observers in the continuous-KR condition were informed that they would receive feedback throughout the training session. Note that the difference between the continuous- and partial-KR conditions was in the number of periods during which KR was provided rather than the information imparted by the feedback.

RESULTS

Correct Detections

Training Phase. The mean percentages of correct detections during both training and transfer are reported in Table 1. For *training*, analysis of variance (ANOVA) confirmed a significant KR effect, $F(2,45)=8.18$, $p<.01$, $\omega^2 =.09$. The effect for the other factors lacked statistical significance. In this and all subsequent analyses proportions were converted to arcsines to normalize the data. The means and standard errors reported are based on the untransformed percentages. Further, in all analyses, Box's correction was employed to adjust for violations of sphericity. Tukey HSD comparisons indicated that the detection scores of observers in the no-KR condition were significantly higher than those of observers who received either form of KR, but that detection scores in the two KR conditions did not differ significantly from one another.

Transfer Phase. During the *transfer* phase observers in the no-KR group attained higher detection scores than those in the KR groups. An ANOVA indicated a significant main effect for KR, $F(2,45)= 12.00$, $p<.001$, $\omega^2 =.08$, for periods, $F(3,152)= 9.02$, $p<.001$, $\omega^2 =.12$, and for the interaction, $F(7,152)=2.04$, $p<.05$, $\omega^2 =.05$. Simple effects tests of periods within each KR condition indicated a significant decline over time for observers in the continuous-KR group, $F(4,53)=5.89$, $p<.01$, $\omega^2 =.20$, and the no-KR group, $F(3,44)=6.18$, $p<.001$, $\omega^2 =.20$. Detection scores for observers in the partial-KR group did not significantly change over time.

<Insert Table 1>

False Alarms

Training Phase. The mean percentages of false alarms during the training and transfer sessions are reported in Table 1. ANOVA revealed significant effects for KR, $F(2,45)= 9.37$, $p<.001$, $\omega^2 =.10$, and for periods, $F(2,88)= 6.75$, $p<.01$, $\omega^2 =.07$. The interaction between these factors was not statistically significant.

Tukey HSD comparisons indicated that participants in the no-KR condition committed significantly more false alarms than those in either KR condition, but scores in the two KR groups did not differ significantly from one another. Across the KR conditions false alarms declined over periods.

Transfer Phase. During *transfer*, observers in the no-KR control group committed substantially more false alarms than those who received KR training. In fact, 41% of all the false alarm scores were zero for the KR groups across the five periods of watch. An ANOVA indicated a significant effect for KR, F (2,45)= 13.16, p<.001,ω^2 =.09, and a significant interaction, F (7,166)= 2.24, p<.05, ω^2 =.04. The period effect was not significant. Simple effects tests of periods within each KR condition indicated that false alarms committed by observers in the no-KR condition declined significantly over the watch period, F(2, 33)=3.52, p<.05, ω^2 =.11. False alarm scores for observers in the two KR conditions did not change significantly over time.

Signal Detection Analysis

The substantial number of zero false alarm rates in the KR groups rendered application of parametric SDT analysis inadvisable due to the problems associated with computation of such indices under these conditions (see Davies & Parasuraman, 1982; Macmillan & Creelman, 2005). To circumvent these concerns Craig (1979) suggested that non-parametric sensitivity indices (specifically, A′) should be used when parametric measures are inappropriate. Work by See et al. (1997) established that among the available, non-parametric indices of response bias, β''_D was the most effective for vigilance. The correct detection and false alarm scores for each participant were therefore used to compute A′ and β''_D.

Perceptual Sensitivity (A′)

Training Phase. An ANOVA of the A′ scores during *training* indicated that sensitivity did not significantly differ across the three KR conditions, and that A′ scores for observers remained stable with time on watch. The interaction between these factors also lacked significance.

Transfer Phase. Sensitivity scores during *transfer* for the three KR conditions are plotted as a function of periods of watch in Figure 1a. An ANOVA indicated a significant effect for KR, F (2,45)= 6.82, p<.01, ω^2 =.05, and a significant effect for period, F (4,104)= 4.20, p<.05, ω^2 =.05. Sensitivity of observers declined over time in all conditions, confirming the classic "vigilance decrement" (See et al.,1995). The interaction was not statistically significant. Tukey HSD tests indicated that observers in the continuous-KR group (M=.85; d=-.78)

were significantly less sensitive than those in the no-KR condition (M=.91). A′ scores for observers in the partial-KR condition (M=.88; d=-.53) did not differ significantly from those in the other two groups.

<Insert Figure 1 about here>

Response Bias (β_D'')

Training Phase. An ANOVA on β''_D scores during *training* indicated a significant effect for KR, F (2,45)= 13.93, p<.001 ω^2 =.15. However, there were no significant differences in β''_D scores across periods, and the interaction was also not statistically significant. Tukey HSD tests revealed that observers in the continuous-KR (M=.94; d=.84) and partial-KR (M=.95; d=.87) groups did not significantly differ from one another, but both groups were significantly more conservative than those in the no-KR condition (M=.54).

Transfer Phase. Mean response bias scores for observers in the three KR conditions during *transfer* are plotted as a function of periods of watch in Figure 1b. Observers who received either form of KR achieved a conservative level of responding early in the watch and remained at that level, while observers who did not receive KR were more lenient in responding and became more conservative over time. ANOVA revealed a significant effect for KR, F (2,45) = 13.21, p<.001, ω^2 =.09, for period, F (4,77) = 7.92, p<.01, ω^2 =.10, and for the interaction, F (8,77)= 6.85, p<.001, ω^2 =.16. Tests for the simple effects of period within each KR condition confirmed that observers in the no-KR group became more conservative over time, F (4,77)= 21.49, p<.01, ω^2 =.51, but that β''_D scores of observers in the two KR groups did not change significantly over time.

Diagnosticity Measures

The PPP and NPP were computed for each participant in each period of watch during both training and transfer sessions. It is important to remember that in vigilance the number of non-signals far exceeds the number of signals, so that opportunities for errors of omission are much rarer than opportunities for false alarms. Thus, in cases where there is a low signal probability and many correct rejections (i.e., the observer adopts a conservative criterion), NPP values would be expected to be artificially high (above 90%). Therefore, any differences in NPP among KR conditions will likely be driven by differences in errors of omission. In contrast, PPP is influenced by frequencies of correct detections and false alarms, and these responses are relatively few in number compared with the total number of events during a vigil.

Positive Predictive Power (PPP)

Training Phase. An ANOVA of the PPP scores revealed a significant period effect, $F(2,83)= 5.50$, $p<.01$, $\omega^2 =.06$, with scores increasing over time (M's for periods 1-3= .62, .68, and .71, respectively). No significant differences among KR groups were observed, and the interaction also lacked statistical significance. *Transfer Phase*. Mean PPP scores for the three groups are displayed as a function of period on watch in Figure 2a. ANOVA revealed a significant effect for KR, $F(2,45)= 6.92$, $p<.01$, $\omega^2 = .05$. No significant differences were observed for period or the KR by period interaction. Tukey HSD tests revealed that the two KR groups did not significantly differ from one another (continuous-KR M=.83; d=.96; partial-KR M=.81; d=.88), but the PPP scores of both these groups exceeded those of the no-KR group (M=.63).

<Insert Figure 2 about here>

Negative Predictive Power (NPP)

Training Phase. An ANOVA revealed a significant effect for KR group, $F(2,45)= 11.69$, $p<.01$, $\omega^2 =.13$. No significant effects were observed for periods or the interaction between these factors. Tukey HSD tests revealed that the no-KR group (M=.98) achieved NPP scores that were significantly higher than those of the continuous-KR (M=.97; d=.72) and partial-KR groups (M=.96; d=.99). The NPP scores of the two KR conditions did not differ significantly from one another. Although these absolute mean differences are small, the effect sizes are large as a result of low within-group variability.

Transfer Phase. Mean NPP scores for the three groups are displayed as a function of period on watch in Figure 2b. Observers in the no-KR condition achieved higher NPP scores during transfer than those in the two KR conditions. ANOVA indicated significant effects for KR, $F(2,45)= 12.20$, $p<.001$, $\omega^2 = .09$, periods, $F(3,141)= 10.93$, $p<.001$, $\omega^2 = .14$, and the interaction, $F(6, 141)= 2.82$, $p<.05$, $\omega^2 = .06$. Simple effects tests of periods within each KR condition indicated that NPP scores significantly declined over the watch period in the no-KR condition, $F(3,141)= 11.07$, $p<.01$, $\omega^2 =.14$, and the continuous-KR condition, $F(3,141)= 4.30$, $p<.01$, $\omega^2=.05$. NPP scores for participants in the partial-KR condition did not change significantly over time.

DISCUSSION

Signal Detection

In this experiment we found no evidence that KR enhanced perceptual sensitivity. This was surprising, given that KR typically enhances signal detection in vigilance (Warm & Jerison, 1984). However KR did increase conservatism in responding relative to a no-KR condition, a finding that has been observed previously in vigilance experiments (See et al., 1997). The increase in conservatism over time on watch in the no-KR condition is also consistent with prior research, and may be due to the growing awareness over the vigil that critical signals are relatively rare (Craig, 1978). The provision of KR likely produces an immediate and explicit awareness of the rarity of signals and so causes observers to set a high criterion very early in the vigil.

There was a tradeoff between PPP and NPP such that the enhanced diagnosticity provided by KR for 'yes' responses (PPP) is achieved at the cost of lower diagnosticity for 'no' responses (NPP). Consideration of both the SDT and diagnosticity metrics indicates that KR increases conservatism and improves the reliability of 'yes' responses, but that this occurs at the expense of lower accuracy in 'no' decisions and lower perceptual sensitivity. That is, KR facilitated learning to avoid false alarms, reflected in enhanced PPP and increased conservatism, but did not facilitate learning to avoid misses, manifested in lower levels of NPP and sensitivity.

The differential efficacy of KR for reducing false alarms versus misses may have resulted in part from

the perceptual discrimination required. As in most vigilance experiments, stimuli were presented on the screen only for a brief duration. These brief exposure times, coupled with the spatial discrimination required, induced gamma motion in that the two vertical lines appeared to 'move' from the center dot when they appeared on the screen (see Figure 1). This apparent motion may have induced leniency in responding which was reduced by the feedback. However, this task feature may have distracted observers in the KR conditions from learning the distances to be discriminated. Apparent motion did not affect vigilance per se, since the no-KR control results replicate those typically observed. Rather, the apparent motion may have interacted with the informational value of each form of KR which subsequently interfered with the learning process in regard to sensitivity.

Evidence that different forms of KR are not equivalent in their informational value or in their performance effects comes from the work of Dittmar, Warm, and Dember (1985). They provided observers with feedback regarding correct detections, false alarms, or misses. While hit and false alarm KR enhanced detection accuracy, miss KR *failed* to enhance observers' detection ability. These effects occurred despite the fact that hit and false alarm KR should, in principle, provide observers with the same informational feedback.

The relative effectiveness of KR for reducing false alarms versus misses may also depend on whether the feedback occurs after overt responding. In most vigilance experiments overt responses were required only when the observer believed a signal was present (for one exception see Parasuraman & Davies, 1976). Non-signal events therefore require the 'response of not responding,' and no KR is provided for correct rejections. It may be that the KR typically used in vigilance is effective only when observers make an overt response and thereby link the feedback to a specific self-generated event. KR that cannot be so associated with a discrete action (miss KR) may therefore be less effective. Indeed, Dittmar et al. (1985) argued that miss KR does not adequately provide sufficient information to be perceptually useful. Requiring overt responses for non-signals may therefore enhance the efficacy of KR especially for improving NPP and sensitivity. Further exploration of this potentiality is required to establish what elements of KR produce these differential performance effects.

Criterion Setting or Diagnosticity?

The results of the SDT analysis might tempt one to conclude that the effects of KR on PPP and NPP are driven solely by criterion setting. However, reduction in responding *per se* does not need to result in a change in a particular direction of either PPP or NPP. This can be seen by the example in Table 2, which shows hit and false alarm rates that produce common levels of response bias but different values of PPP and NPP, and a comparable example analysis in which A′ is held constant. The diagnosticity measures are determined by the relative frequencies of correct and incorrect decisions, rather than the absolute level of 'yes' responses. Therefore, both the operator's perceptual ability and level of responding can influence diagnosticity.

<Insert Table 2 about here>

KR Schedule

One goal for this study was to investigate the possibility that observers who receive partial-KR training perform better during a test vigil than those who experience continuous-KR training. This hypothesis was not supported, an outcome that differs from that obtained by Warm, Hagner, and Meyer (1971). They reported that partial-KR training facilitated performance during a test vigil. However, Warm and his colleagues evaluated *reaction time* while this study evaluated *response accuracy*. These aspects of performance are thought to reflect different perceptual processes (Santee & Egeth, 1982). Reaction time to easily detectable stimuli, such as those employed by Warm and his colleagues, likely reflects resource-limited processing. This in turn may be influenced by motivational processes. By contrast, the present detection task required a much more difficult discrimination and is much more likely to be data-limited and thus not influenced by increased effort (c.f., Norman & Bobrow, 1975). Partial-KR may be more effective than continuous-KR when performance is in the resource-limited range as participants are able to modulate their effort. They may reduce their effort during the KR portions of the experiment and increase effort during the portions in which no feedback is provided. In the data-limited range, effort modulation will not influence performance, rendering the two forms of KR equally efficacious. This interpretation is consistent with an effort-regulation model in which effort level is modulated according to the level of demand (see Hockey,

1997). Modulation of effort may be effective when the task is resource-limited, but in the data-limited range such regulation is less flexible and therefore less beneficial.

SDT and PPP/NPP

The effects of KR and time on task in this study were contingent upon the dependent measures employed. A′ showed a decline over time and a negative effect of KR during transfer, and β''_D showed a KR effect during training and an interaction during transfer. PPP showed no time effects during either session, but did show a facilitative effect for KR during transfer. The NPP analysis indicated a debilitative effect of KR during training and an interaction between KR and time during transfer. Thus, the two sets of measures did not produce the same patterns of outcome. The question is not which set of metrics is better, but rather how the metrics may be used together to provide a more complete performance portraiture. Sensitivity provides information regarding stimulus discriminability, while response bias reflects the decision criterion. These measures correspond more directly to psychological processes although they do not imply specific mechanisms. Indeed, Swets (1977) argued that there is more to vigilance than discrimination and criterion setting, and that SDT cannot provide a total explanation for its performance effects. Further, sensitivity measures often reflect more than pure sensory processing ability (Pastore, Crawley, Berens, & Skelly, 2003). So, while SDT indices may tend to reflect psychological processes they do not, for this reason alone, supersede other useful measures. The PPP/NPP approach provides a direct assessment of the accuracy of yes/no responses given particular levels of sensitivity and bias, and thereby indicates how often the individual's decision will be correct. Given that poor diagnostic decisions can occur when signal base rates are low (Parasuraman et al., 1997), measures that directly reflect diagnostic power can be especially useful in evaluating real-world detection systems.

Practical Applications

The Transportation Security Administration (TSA) has placed over 1 billion dollars of detection equipment into operation at over 400 airports within the United States (Airports Staff, 2001). The efficiency of such systems depends crucially upon the capability of operators who use them. To enhance operator

efficiency technologies such as the Threat Image Projection System (TIP; Neiderman & Fobes, 1997) overlay computer-generated images on to inspected items to provide feedback for performance improvement. Here we demonstrate that the role of feedback in vigilance can be interpreted differently, contingent on the nature of the dependent measure employed. The relative importance of diagnosticity for signal and non-signal events is therefore critical in evaluating the effectiveness of KR manipulations as training strategies for monitoring tasks. Our results show that KR effectiveness depends in part on whether emphasis in training is on false-alarm reduction or on the reduction of missed signals. For the latter, it might be advisable to require operators to make overt 'no' responses during training to attenuate the decline in the accuracy.

Aviation security issues are only one example among many domains in which training for vigilance has practical utility. Another obvious realm of concern is detection and decision making in medicine. While PPP and NPP are relatively new to human factors professionals, these measures are often used by health professionals to whom the human factors community is now reaching out (Bogner, 1994). Therefore, we must assure those professionals with whom we wish to interact that we can embrace and integrate their measurement approaches with those which we are already comfortable and familiar. This study is the first direct comparison between PPP/NPP with the classic measures of signal detection in sustained attention. Our results have shown that PPP/NPP provide another window on performance which is especially useful when false alarm rate is naturally low. We hope that others will see the advantage of these differing methods and, given the simple calculations needed for their derivation, include PPP and NPP in their own efforts to understand operators' capabilities to detect environmental signals.

ACKNOWLEDGEMENT

Completion of this work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. P.A. Hancock, *Principal Investigator.* The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant. We are also grateful to Eric Warm, M.D. for helpful discussions regarding the application of diagnosticity measures in medicine.

REFERENCES

Airports Staff. (2001, November 15). Airports: S.1447, Aviation and Transportation Security Act. *Aviation Week and Space Technology*. Retrieved January 29, 2004 from http://www.aviationnow.com

Bogner, M.S. (1994).(Ed.). *Human error in medicine*. Hillsdale, NJ: Erlbaum.

Craig, A. (1978). Is the vigilance decrement simply a response toward probability matching? *Human Factors*, *20,* 441-446.

Craig, A. (1979). Nonparametric measures of sensory efficiency for sustained monitoring tasks. *Human Factors, 21*, 69-78.

Davies, D.R., & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Dittmar, M.L., Warm, J.S., & Dember, W.N. (1985). Effects of knowledge of results on performance in successive and simultaneous vigilance tasks: A signal detection analysis. In: R.E. Eberts and C.G. Eberts

(Eds.), *Trends in ergonomics/human factors II.* (pp. 195-202). Amsterdam: Elsevier.

Getty, D.J., Swets, J.A., Pickett, R.M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1,* 19-33.

Hancock, P.A., & Hart, S.G. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in Design, 10,* 6-16.

Harris, D. (2002). How to really improve airport security. *Ergonomics in Design*, *10*, 17-22.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*, 73-93.

Linton, C.S. (1996). General internal medicine. In: U.B.S. Prakash (Ed.), *Mayo Internal Medicine Board Review 1996-1997.* (pp. 333-352). Rochester, MN: Mayo Foundation for Medical Education and Research.

Macmillan, N.A., & Creelman, C.D. (2005). *Detection theory: A user's guide*. (2nd Edition). Mahwah, NJ: Erlbaum.

McCormack, P.D., Binding, F.R.S., & McElerhan, W.G. (1963). Effects on reaction time of partial knowledge of results of performance. *Perceptual & Motor Skills, 17*, 279-281.

Neiderman, E.C. & Fobes, J.L. (1997). Operational test and evaluation of human factors interventions for airport security screeners. *Proceedings of the Human Factors and Ergonomics Society*, *41*, 1094-1097. Santa Monica, CA: Human Factors and Ergonomics Society.

Norman, D.A., & Bobrow, D.G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

Parasuraman, R., & Davies, D.R. (1976). Decision theory analysis of response latencies in vigilance. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 578-590.

Parasuraman, R., Hancock, P.A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision warning systems. *Ergonomics, 39*, 390-399.

Pastore, R.E., Crawley, E.J., Berens, M.S., & Skelly, M.A. (2003). "Nonparametric" A' and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review, 10*, 556-569.

Salmoni, A.W., Schmidt, R.A., & Walter, C.B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin, 95*, 355-386.

Santee, J.L., & Egeth, H.E. (1982). Do reaction time and accuracy measure the same aspects of letter recognition? *Journal of Experimental Psychology: Human Perception and Performance, 8*, 489-501.

See, J.E., Howe, S.R., Warm, J.S., & Dember, W.N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117*, 230-249.

See, J.E., Warm, J.S., Dember, W.N., & Howe, S.R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors, 39*, 14-29.

Swets, J.A. (1977). Signal detection theory applied to vigilance. In: R.R. Mackie (Ed.), *Vigilance: Operational performance and physiological correlates.* (pp. 705-718). New York: Plenum Press.

Szalma, J.L, Miller, L.C., Hitchcock, E.M., Warm, J.S., & Dember, W.N. (1998). Intraclass and interclass transfer of training for vigilance. In: M.W. Scerbo and M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 183-187). Mahwah, NJ: Erlbaum.

Warm, J.S. (1993). Vigilance and target detection. In: B.M. Huey and C.D. Wickens, (Eds.), *Workload transition: Implications for individual and team performance.* (pp.139-170). Washington, DC: National Academy Press.

Warm, J.S., Hagner, G.L., & Meyer, D. (1971). The partial reinforcement effect in a vigilance task. *Perceptual & Motor Skills, 32*, 987-993.

Warm, J.S., & Jerison, H.J. (1984). The psychophysics of vigilance. In: J.S. Warm (Ed.), *Sustained attention in human performance.* (pp. 15-59). Chichester, United Kingdom: Wiley.

Table 1. Mean percentage of correct detections and false alarms during training and transfer for three KR conditions

	Period on Watch						
Correct Detections	1	2	3	4	5	Mean	Cohen's d
Training							
No KR	67 (7)	70 (4)	67 (6)			68 (4)	
Continuous KR	62 (5)	52 (5)	50 (3)			54 (4)	.88
Partial KR	55 (5)	44 (5)	44 (4)			48 (4)	1.25
Mean	61 (3)	56 (3)	54 (3)				
Transfer							
No KR	77 (5)	74 (5)	65 (6)	64 (5)	63 (4)	69 (4)	
Continuous KR	61 (.04)	45 (4)	46 (5)	41 (5)	38 (4)	46 (3)	-1.53
Partial KR	62 (4)	50 (4)	55 (4)	54 (4)	58 (4)	56 (2)	-.98
Mean	67 (3)	56 (3)	55 (3)	53 (3)	53 (2)		
False Alarms							
Training							
No KR	4.4 (.7)	4.6 (.7)	3.5 (.7)			4.1 (.4)	
Continuous KR	2.5 (.4)	1.0 (.2)	1.2 (.2)			1.5 (.4)	.94
Partial KR	2.2 (.4)	1.2 (.2)	1.5 (.3)			1.6 (.4)	.90
Mean	3.0 (.3)	2.3 (.3)	2.0 (.3)				
Transfer							
No KR	4.6 (.8)	4.6 (.7)	3.7 (.7)	3.6 (.7)	3.3 (.6)	4.0 (.6)	
Continuous KR	1.2 (.3)	.5 (.2)	.7 (.2)	.9 (.3)	.7 (.2)	.8 (.1)	-1.58
Partial KR	1.2 (.3)	.9 (.3)	.9 (.2)	.9 (.3)	1.7 (.5)	1.1 (.2)	-1.39
Mean	2.3 (.3)	2.0 (.3)	1.8 (.3)	1.8 (.3)	1.9 (.3)		

Note. Values in parentheses are standard errors. Cohen's d was computed by comparing each experimental group to the No KR control condition.

Table 2. Diagnosticity measures corresponding to hit/false alarm pairs with the same response bias level.

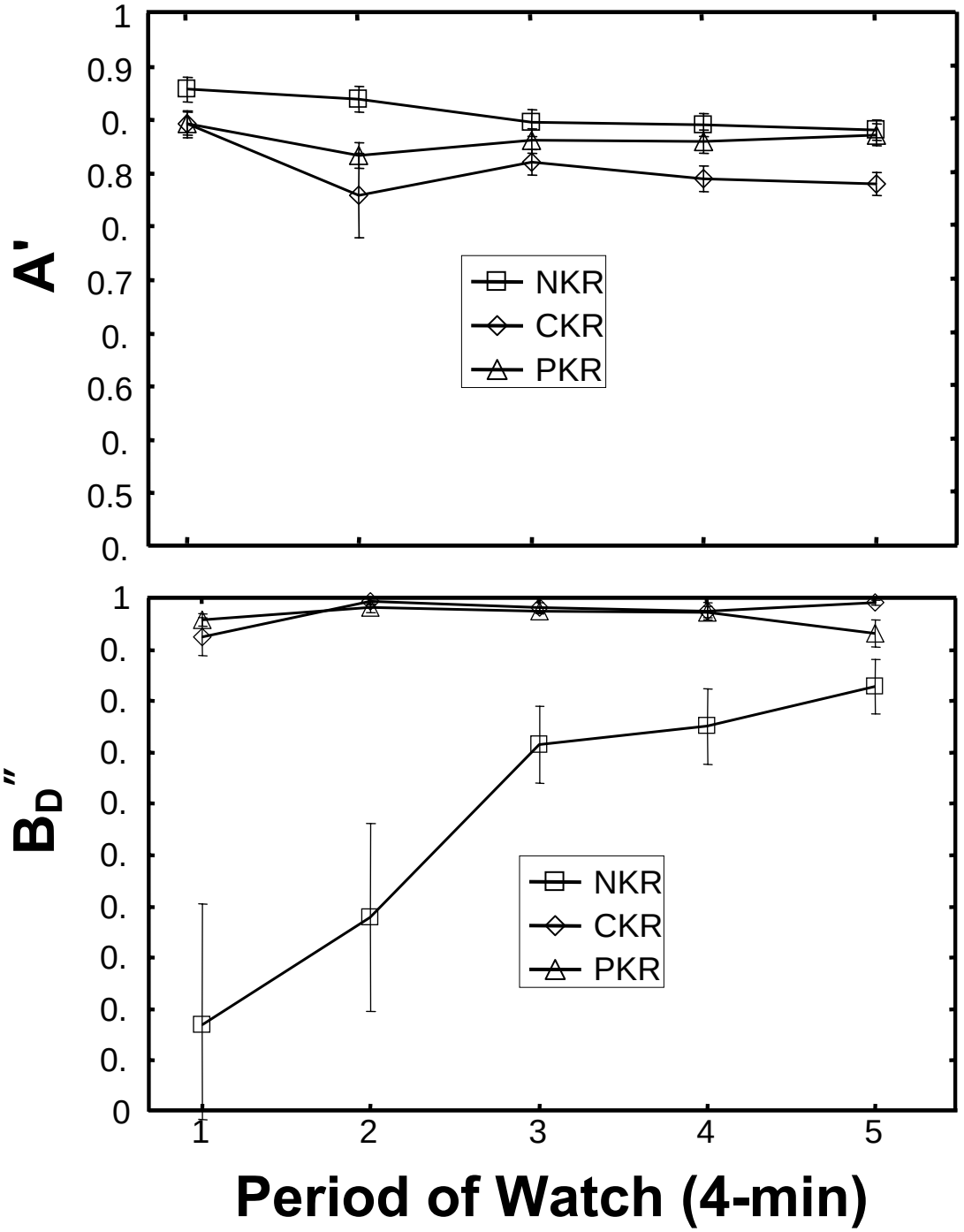
A'	B_D''	p(H)	p(FA)	PPP	NPP
0.50	0.99	0.077	0.053	0.09	0.93
0.50	0.99	0.120	0.031	0.22	0.94
0.51	0.99	0.178	0.017	0.43	0.94
0.52	0.99	0.250	0.009	0.66	0.95
0.59	0.99	0.500	0.004	0.89	0.97
0.70	0.99	0.688	0.002	0.96	0.98
0.80	0.99	0.812	8.93×10^{-4}	0.98	0.99
0.86	0.99	0.875	8.93×10^{-4}	0.99	0.99
0.96	0.99	0.969	1.19×10^{-4}	1.00	1.00
0.71	0.08	0.838	0.143	0.30	0.99
0.71	0.27	0.812	0.116	0.33	0.98
0.71	0.35	0.800	0.107	0.35	0.98
0.71	0.60	0.762	0.071	0.43	0.98
0.71	0.66	0.762	0.061	0.47	0.98
0.71	0.76	0.738	0.046	0.54	0.98
0.71	0.84	0.736	0.031	0.63	0.98
0.71	0.95	0.712	0.010	0.84	0.98
0.71	1.00	0.700	8.93×10^{-6}	1.00	0.98

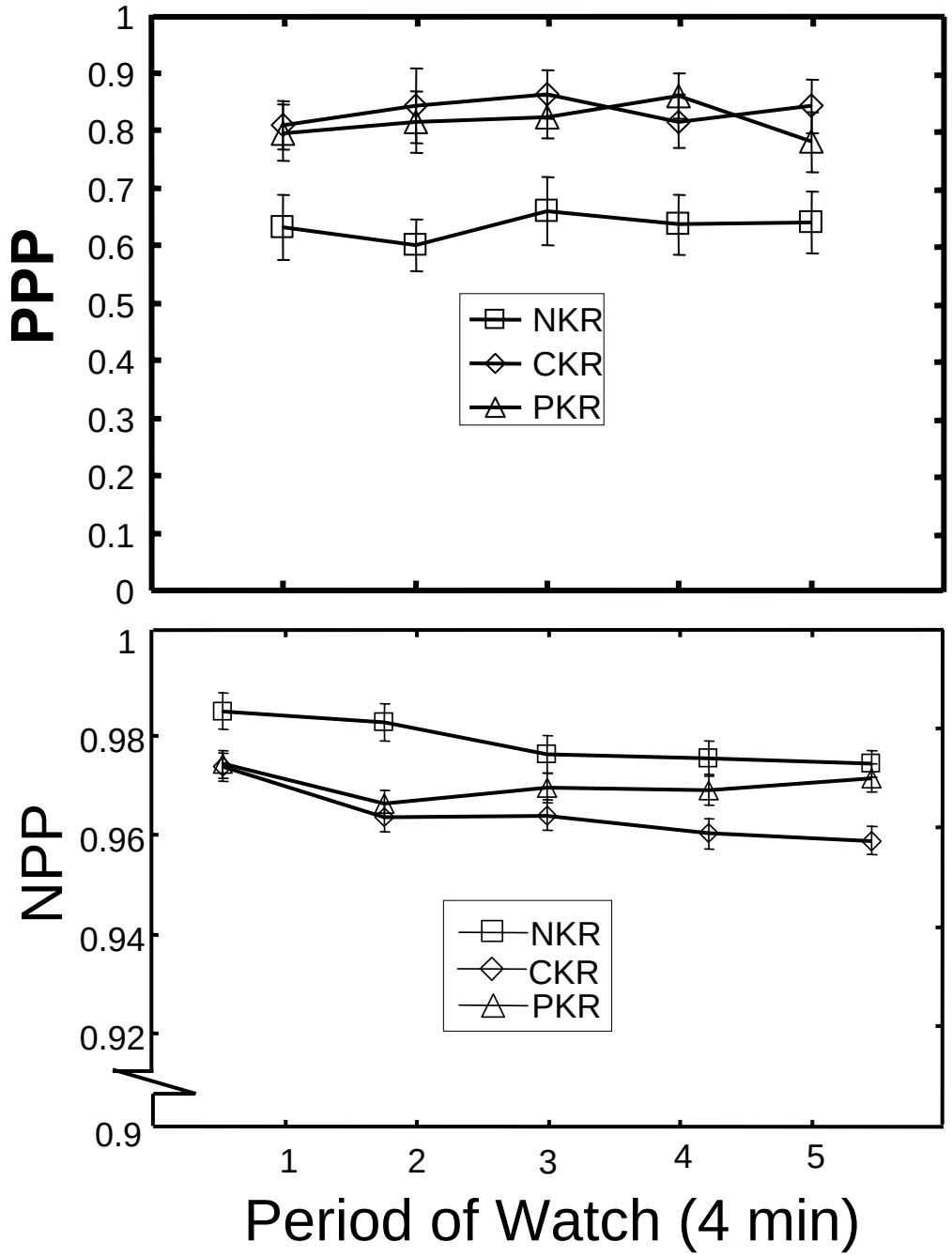
Note. Values in this table assume base rates used in this study (signal rate of .067 and a non-signal rate of .933). PPP=positive predictive power; NPP=negative predictive power.

Figure Captions

Figure 1. (a) Mean sensitivity (A') scores as a function of periods of watch during transfer. (b) Mean response bias (β_D'') scores as a function of periods of watch during transfer. *Note*. CKR = continuous KR; PKR = partial KR; NKR = no KR. Error bars are standard errors.

Figure 2. (a) Mean positive predictive power (PPP) as a function of periods of watch during transfer. (b) Mean negative predictive power (NPP) as a function of periods of watch during transfer. *Note*. CKR = continuous KR; PKR = partial KR; NKR = no KR. Error bars are standard errors.


A'
1
0.9
0.8
0.7
0.6
0.5
NKR
CKR
PKR
B"D
1
0
1
2
3
4
5
Period of Watch (4-min)



Biographies

James L. Szalma is a Research Scientist at the University of Central Florida. He received his Ph.D. in experimental psychology/human factors from the University of Cincinnati in 1999.

Peter A. Hancock is the Provost Distinguished Professor of Psychology at the University of Central Florida, Orlando, FL. He received his Ph.D. in motor performance from the University of Illinois at Urbana-Champaign in 1983.

Joel S. Warm is Professor of Psychology and director of graduate studies in psychology at the University of Cincinnati. He received his Ph.D. in experimental psychology from the University of Alabama in 1966.

William N. Dember is Dean Emeritus and Professor Emeritus at the University of Cincinnati. He received his Ph.D. in experimental psychology from the University of Michigan in 1955.

Kelley S. Parsons is a Graduate Student in the doctoral program in experimental psychology/human factors at the University of Cincinnati. She received an M.A. in experimental psychology/human factors from the University of Cincinnati in 2001.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

British Journal of Psychology (2006), 97, 115–135

The British Psychological Society

www.bpsjournals.co.uk

Training for vigilance: The effect of knowledge of results format and dispositional optimism and pessimism on performance and stress

James L. Szalma[1]*, Peter A. Hancock[1], William N. Dember[2] and Joel S. Warm[2]

[1]Department of Psychology and Institute for Simulation and Training, University of Central Florida, USA

[2]Department of Psychology, University of Cincinnati, USA

This study investigated the impact of knowledge of results (KR) format on the performance and stress associated with a vigilance task. Also examined was the effect of the interaction of KR-format (Hit-KR, False Alarm-KR, Miss-KR, and a composite of all three formats) with dispositional optimism and pessimism on performance outcome and reported stress state. Hypotheses based upon a theory of feedback intervention were tested. KR regarding correct detections and the composite-KR (KR regarding correct detections, false alarms, and missed signals) enhanced perceptual sensitivity. However, False Alarm-KR and Miss-KR did not. Contrary to expectations based on the theory, performance was unrelated to the traits across all KR conditions. However, the effects of KR-format on self-reports of stress depended on the individual's level of pessimism and optimism. In addition, KR format and personality affected the multiple dimensions of stress state in different ways.

Tasks demanding vigilance or sustained attention require observers to monitor displays for the appearance of infrequent signals for action over prolonged periods of time. The central issue, in almost 60 years of vigilance research, has been the decline in performance with time on watch termed the vigilance decrement (See, Howe, Warm, & Dember, 1995). Although vigilance has been and remains an important part of many contemporary work environments, it has assumed significantly greater importance with the introduction of many 9/11-mandated security systems such as by the Transportation Security Administration's (TSA) need to screen 100% of all baggage at all US airports (Harris, 2002 see also Hancock & Hart, 2002). The advent of these, and many similar, security requirements has increased the need for effective training procedures for system operators. In vigilance, training has traditionally employed feedback in the form of knowledge of results (KR) regarding an operator's accuracy expressed in terms of

* *Correspondence should be addressed to James L. Szalma, MIT[2] Laboratory, 3100 Technology Parkway, Ste 333, University of Central Florida, Orlando, FL 32826, USA (e-mail: jszalma@mail.ucf.edu).*

DOI:10.1348/000712605X62768

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

correct detections, false alarms, or missed signals. KR has been shown to improve performance on sustained attention tasks (for a review, see Warm & Jerison, 1984). However, there is experimental evidence that the form of KR can impact its effectiveness. For example, Dittmar, Warm, and Dember (1985) asked participants to monitor increments in a vertical line presented on a video display terminal (VDT) and provided observers with feedback, in the form of a 1,000 Hz tone, regarding their correct detections, false alarms or missed signals. They observed that hit- and false alarm-KR enhanced perceptual sensitivity (A') and attenuated the vigilance decrement, but miss-KR had no such effects. They explained these findings in terms of the information provided by each KR format. Hit- and false alarm-KR provide information regarding responses to specific stimuli but miss-KR is not linked to the observer's noticing of any particular characteristic of a signal event or to the observer's subsequent response.

Although Dittmar and her colleagues (1985) reported performance effects of KR format, they did not examine the stress effects associated with the use of KR. Several experiments have reported that vigilance tasks are stressful. Observers report feeling more boredom, strain, irritation, distress and less task engaged after the task than prior to its start (e.g. see Szalma *et al.*, 2004; Warm, 1993). In addition, research has indicated that manipulations of task characteristics that influence performance also influence individual subjective state (e.g. Szalma *et al.*, 2004; Temple *et al.*, 2000; see also Warm, Dember, & Hancock, 1996). Therefore, it is feasible that the format in which KR is provided to observers can influence their stress state in addition to influencing their performance. Given that Dittmar *et al.* observed performance differences as a function of KR format, and the associations between performance and subjective state (Warm *et al.*, 1996), one might expect that hit- and false alarm-KR would be associated with reduced stress symptoms and that miss-KR would be associated with greater stress, relative to a no-KR control condition. That is, the effect of KR format on stress should mirror the effects previously reported by Dittmar and her colleagues with regard to performance. Previous studies using a composite format have shown that used together the three forms of KR enhance performance (e.g. Szalma, Hancock, Warm, Dember, & Parsons, in press; Szalma, Miller, Hitchcock, Warm, & Dember, 1998), but these studies did not examine the effects of composite KR on observers' stress states. It may be that providing all three forms of KR will show the greatest performance enhancement and reduction in stress, since all information regarding detection accuracy is explicitly provided. One goal for the current experiment was to test these propositions.

Individual differences in performance and stress response: Feedback intervention theory (FIT)

From existing evidence, it appears highly likely that the effects of KR format on performance and stress associated with vigilance will depend upon the characteristics of the individual. Indeed, it has been long recognized that general individual differences among observers can influence performance in vigilance (see Berch & Kanter, 1984; Davies & Parasuraman, 1982). Recent evidence has confirmed that responses to the stress of sustained attention also depend on characteristics of the individual observers (e.g. Helton, Dember, Warm, & Matthews, 1999; Szalma, 2002a). Such findings are consistent with transactional models of stress that emphasize the importance of appraisal processes in stress response (Lazarus & Folkman, 1984; Matthews, 2001).

An individual differences variable that has recently emerged as an influential factor on stress responses to sustained attention is dispositional pessimism/optimism.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Helton *et al.* (1999) investigated the effect of false failure feedback on the performance and stress of both pessimists and optimists in a vigilance task. Although false failure feedback did not significantly impact performance, individuals high in pessimism and low in optimism exhibited steeper vigilance decrements across conditions relative to individuals low in pessimism and high in optimism. Helton and his colleagues also observed that pessimistic individuals reported changes in stress-related parameters such as greater tension, lower hedonic tone (i.e. a more negative mood state) and less confidence relative to optimists. These findings suggest that the high stress associated with vigilance may be moderated by an observer's relative level of pessimism/optimism. Further evidence for the influence of pessimism and optimism on observers' subjective stress states was reported by Szalma (2002a) who found that higher degrees of pessimism were associated with greater *distress* and lower *task engagement* after a vigilance task. Szalma also reported that pessimism predicted greater reliance on emotion-focused coping, consistent with prior research indicating use of this coping strategy by pessimistic individuals. Optimism, in contrast, was less predictive of stress and coping.

With regard to individual differences in the effectiveness of KR for vigilance, predictions can be made based upon a theory of feedback intervention (FIT), derived from a meta-analytic study of the feedback intervention literature (Kluger & DeNisi, 1996). FIT serves as a general framework for understanding the processes that influence feedback effectiveness. Drawing on control theory (see Carver & Scheier, 1981, 1998), FIT assumes that behaviour is regulated by comparison of feedback to goals or standards that are hierarchically organized. Two further assumptions of this theory are that (1) only those feedback interventions that receive attention can influence performance, and (2) feedback interventions have the capacity to alter the locus of attention within the goal hierarchy. Although hierarchies can vary in complexity, Kluger and DeNisi describe three general levels relevant for feedback effects. At the lowest level are *task-learning processes* consisting of 'hypotheses' for improving performance that individuals test by utilizing feedback. Attention to this level of the hierarchy constitutes attention to the details of focal task performance. These processes are regulated by the next highest level in the hierarchy, *task-motivation processes*, which serve as standards or comparators for the task-learning processes. At the highest level are *meta-task processes* involving the self and general but highly salient goals. Specifically, *meta-task processes* relate task outcomes to higher-level goals that are important to the individual. Kluger and DeNisi argued that most affective reactions to feedback interventions, when they occur, result from evaluation of feedback with respect to salient self-goals specified at the meta-task level.

According to FIT, feedback is effective in improving performance only to the extent that it serves to focus attention on the task (i.e. *task-motivation* and *task-learning processes*). Thus, an FIT interpretation of the results of Dittmar and her colleagues (1985) would be that the hit- and false alarm-KR directed attention to task-learning processes (e.g. learning the nature of a signal/non-signal) and perhaps task-motivational processes (e.g. see Warm, Kanfer, Kuwada, & Clark, 1972; Warm, Epps, & Ferguson, 1974), while miss-KR failed to adequately direct attention to relevant task characteristics. Miss-KR may fail to direct attention appropriately because it is not linked to a specific, overt response from the observer, so that observers are unable to use the feedback to associate their response to the stimulus characteristics (see Szalma *et al.*, in press). Moreover, one of the propositions of FIT (see Proposition 5, Kluger & DeNisi, 1996, p. 269) is that feedback cues that match salient personal goals associated with a given personality trait will direct attention to meta-task processes and thereby debilitate performance. This may subsequently induce stress effects in such individuals.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Kluger and DeNisi suggested a general mechanism for these effects whereby individuals' traits influence their ability to redirect their attention to the details of the task in response to feedback. They proposed that this occurs via an inhibition mechanism that protects working memory space from competing demands imposed by the goal hierarchy, and that traits such as the tendency to blame oneself for failures may exert debilitative effects on feedback by interfering with this inhibition mechanism. Hence, individual differences relevant to an observer's response to feedback may influence the effectiveness of training with feedback interventions such as KR.

Given the propositions of Kluger and DeNisi (1996), traits such as pessimism and optimism, which affect general expectancies regarding performance and affective responses to performance success and failure (see Chang, 2002), should influence the effectiveness of KR in improving vigilance. Specifically, the negative outcome expectancies associated with pessimism may serve to interfere with these inhibitory processes and thereby reduce KR effectiveness, particularly for 'negative' KR (i.e. false alarm-KR and miss-KR), since individuals high in pessimism tend to appraise events more negatively than those low in pessimism. Such feedback may, in Kluger and DeNisi's terms, 'release' the pressure on the regulatory system to eliminate the feedback-standard discrepancy at the task level. Attention is then redirected towards the self and managing the negative affect associated with pessimistic appraisals of the task. In contrast, the more positive expectations of individuals high in optimism would result in feedback directing their attention towards more task-based processes, thereby improving performance and reducing stress symptoms. These effects may be weak, however, since optimism has been shown to be a weaker absolute predictor of other measures than pessimism (Szalma, 2002a; see also Hummer, Dember, Melton, & Schefft, 1992). The results of Helton *et al.* (1999) concerning false failure feedback sit contrary to these hypotheses, as failure feedback would be expected to impair performance of pessimists by inducing negative appraisals and directing their attention towards dealing with their subsequent negative affect. However, the absence of significant KR effects in their study may be due to the fact that feedback was not veridical and was provided at the end of a block of trials (i.e. summary KR). Veridical feedback provided after each response (or failure to respond when appropriate) provides many more opportunities for the personality traits to differentially exert their influence on performance and stress state across feedback conditions.

The present study was thus also designed to test these propositions by examining the impact of individual differences in optimism and pessimism on performance and stress in vigilance, and the effectiveness of KR in facilitating performance and reducing stress. Specifically, individuals higher in pessimism should exhibit greater performance impairment and report higher levels of stress symptoms after the vigil. Based on FIT, it is hypothesized that veridical feedback emphasizing failure immediately following an error (i.e. miss- or false alarm-KR) should debilitate performance and increase stress for observers higher in pessimism (due to a shift in focus from the task towards the self) relative to a no-KR control condition. In the composite condition, the effects of negative KR may be offset by the provision of hit-KR. Thus, participant disposition should interact with feedback type, such that performance and stress effects associated with high pessimism should manifest themselves primarily in the context of false alarm- and miss-KR. In contrast, individuals high in optimism may respond positively to hit-KR, although the findings of Szalma (2002a) and previous research on optimism and pessimism (Hummer *et al.*, 1992) indicate that these effects may be weaker than those associated with pessimism.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Experimental method

Experimental participants

Participants were 55 male and 55 female undergraduates from the University of Cincinnati who took part in this experiment in exchange for course credit. Participants ranged from 18 to 30 years old, with a mean of 20.0 years. All participants had normal or corrected-to-normal vision and were free of any known hearing impairments.

Feedback conditions and task characteristics

Of the participants, 22 were assigned at random to each of the following five groups, with the only restriction being that the groups were equated for gender: hit-KR – observers received feedback regarding correct detections; false alarm-KR – observers received feedback when they made errors of commission; miss-KR – observers received feedback when a signal appeared on the screen and they failed to respond to it; composite-KR – observers received all three forms of feedback; no-KR – observers received no feedback regarding performance quality and acted as the comparison control group. The task required monitoring a simulated chemical process display presented on a VDT for patterns of deviation that would lead to an uncontrolled reaction (see Fig. 1). The display consisted of Cartesian coordinates (formed by 3.2 cm green line segments) representing

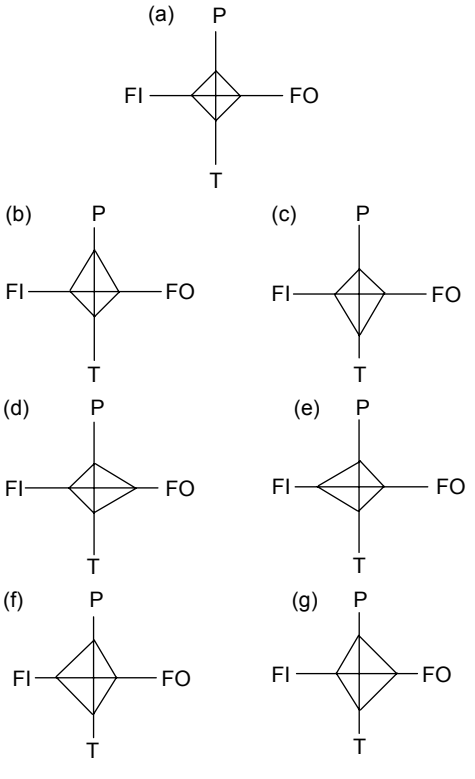


Figure 1. Stimuli employed in the current experiment. Displays (a) through (e) are 'neutral events', requiring no overt response from the observer. Displays (f) and (g) are 'critical signals', defined by simultaneous changes on both axes.

Note. P = pressure; *T* = temperature; FI = flow rate in; FO = flow rate out.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

the state of four system variables, each of which was represented as a point on one of the axes. These points were connected via green line segments, which formed a polygon that changed shape according to changes in system variables. Representative stimuli are displayed in Fig. 1. The variables to be monitored were the pressure, temperature and the flow rates in and out of the 'vat' of chemicals. Neutral events, which required no overt response from the observer, were cases in which these variables were in balance (see Display (a) in Fig. 1) or when their levels varied along one axis (e.g. pressure/temperature; see Displays (b) and (c) in Fig. 1) but not the other (e.g. flow rate; see Displays (d) and (e) in Fig. 1). There were five such events. Under one of these conditions, the four vertices along the axes (each representing the level of a system variable) were 0.8 cm from the origin, resulting in a square-shaped polygon. For the other four neutral events, one of the points moved to 1.2 cm from the origin, resulting in four different orientations of a 'kite' shape. Critical signals for detection were cases in which the flow rate-in did not equal the flow rate-out and *both* temperature and pressure increased. Specifically, critical signals consisted of cases in which (1) the temperature, pressure and flow rate-in points were moved from 0.8 cm from the origin to 1.2 cm from the origin; or (2) the temperature, pressure and flow rate-out points were moved from the origin to 1.2 cm from the origin. Since critical and non-critical events were defined in terms of combinations of variable levels, the state of the overall system was conveyed via the shape of the polygon. Thus, neutral events were cases in which the figure displayed was either 'square-' or 'kite-' shaped, while critical signals were cases in which the quadrangle was not diamond- or kite- shaped.

Displays of this type are referred to as 'object' or 'volumetric' displays, and have been successfully employed by human factors researchers in the design of displays in the workplace (see Bennett & Flach, 1992, for a review; see also Hancock & Szalma, 2003). An advantage of these displays is that they allow observers to assess the state of a multidimensional system by inspection of a geometric shape rather than by inspection of several separate gauges, and recent evidence indicates that such displays can attenuate performance decrement, but that they do not reduce the stress of sustained attention (Szalma, 2002b). Note that the display employed in this study was not intended to simulate a specific real-world process, but rather to simulate the kind of displays that have been used to simplify the inspection of complex, multidimensional systems (Bennett, Nagy, & Flach, 1997).

Questionnaires

Observers' stress states were assessed with the Dundee Stress State Questionnaire (DSSQ, version 2.0; Matthews *et al.*, 1999; see also Matthews *et al.*, 2002), which is a multidimensional, self-report instrument for assessing transient mood states. Specifically, this version of the DSSQ comprised 10 factor-analytically determined scales. These scales have been subjected to a secondary factor analysis, yielding three dimensions associated with stress state, *distress*, *task engagement* and *worry*. Thus, increases in stress are associated with increased scores on the *distress* and *worry* scales and/or a decrease on the *task engagement* scale. Example items for *distress* include ratings of the degree to which adjectives such as 'tense', 'relaxed', 'happy' or 'contented' describe the individual's current subjective state. Examples of items associated with *task engagement* include subjective state ratings of adjectives such as 'energetic'. 'alert' or 'vigorous', as well as ratings of statements such as 'I expect the content of the task will be interesting' and 'I am committed to attaining my performance

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

goals'. Example items associated with the *worry* scale include ratings of statements such as 'I'm reflecting about myself' and 'I thought about the difficulty of the problems'.

Optimism and pessimism were assessed using the optimism/pessimism instrument (OPI) developed by Dember, Martin, Hummer, Howe, and Melton (1989). This instrument consists of a 56-item (18 items indicating optimism, 18 indicating pessimism and 20 filler items) regarding the individual's attitudes and expectations. Scores on the OPI range from 18 to 72. Several studies using the OPI have indicated that optimism and pessimism are not simply polar opposites, but are partially independent (for a review, see Dember, 2002). The OPI was administered pre-vigil, and the DSSQ was administered both pre- and post-vigil.

Procedure

Observers participated in a 30-minute vigil divided into six continuous 5-minute periods on watch. The display was presented for 500 ms every 1.5 seconds (30 events per minute), and critical signals for detection occurred twice/min (signal probability = .066) within each period of watch. KR was provided by a computer using a pre-recorded female voice to announce correct detections ('correct'), missed signals ('miss'), and errors of commission ('false alarm'). In the no-KR condition, the voice announced 'saved' after each response in order to control for added auditory stimulation in the KR-conditions. Previous experiments using the female voice to deliver feedback found that this form of KR effectively enhanced performance (Szalma *et al.*, 1998, 2004).

Upon entering the laboratory, participants surrendered their wristwatches and cell phones, read a brief description of the experiment and signed a consent form. Participants completed the OPI, followed by the pre-task version of the DSSQ. The experimenter then read the instructions for the sustained attention task, which were simultaneously presented on the VDT. Observers had no prior knowledge of the length of the vigil, other than that it would not exceed 90 minutes. At the end of the vigil, participants completed the post-task version of the DSSQ and were then debriefed.

Experimental results

Performance measures

Vigilance performance data are usually analysed using signal detection theory (SDT) indices or their base component measures, hit and false-alarm rates. However, in the present experiment, as in many other vigilance experiments, a substantial proportion (37%) of the false alarm scores were zero across participants, periods and KR group (e.g. Szalma *et al.*, 2004, in press; see also Davies & Parasuraman, 1982). Computation of parametric SDT parameters from zero false alarm scores requires adjustments to the data that can introduce error into the estimates (Craig, 1977; Davies & Parasuraman, 1982; Macmillan & Creelman, 2005). The difficulty in interpreting parametric SDT indices under these conditions led Craig (1979) to argue against the uncritical application of these measures to vigilance data. He suggested that nonparametric sensitivity indices (specifically, A') be used when parametric measures are inappropriate. Subsequent research established that among the available, non-parametric indices of response bias, β''_d (Donaldson, 1992) was the most effective for vigilance, as this measure remained sensitive to non-perceptual manipulations and it is effective in estimating response bias at chance performance levels (See, Warm, Dember, & Howe, 1997). One advantage of non-parametric indices is that they are defined for cases in which there are many

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

instances of zero false-alarm rates, as in the present experiment. Correct detections and false alarms for each observer were therefore used to compute SDT measures A′ and β''_d for each period of watch. A 5 (group) by 6 (period) ANOVA was computed for each of these respective performance measure. Mean sensitivity scores for observers in each KR condition are plotted as a function of period of watch in Fig. 2.

Performance remained relatively stable across the watch period, and observers who received any form of KR achieved higher A′ scores than those who did not receive KR. The ANOVA confirmed a significant effect for KR, $F(4, 105) = 3.85$, $p < .01$, $\omega^2 = .02$, and a significant effect for period on watch, $F(3, 318) = 7.49$, $p < .001$, $\omega^2 = .05$. The latter effect is probably a reflection of the fluctuation in performance over the course of the vigil (*M*s for Periods 1 through 6 were 0.87, 0.88, 0.85, 0.90, 0.86 and 0.88, respectively). The interaction between these factors lacked significance ($p > .05$). For all analyses involving repeated measures, Box's epsilon was used to adjust the degrees of freedom for violations of sphericity (see Maxwell & Delaney, 2004). Tukey HSD tests for the KR conditions revealed that relative to those in the no-KR control group ($M = 0.81$), observers who received composite-KR ($M = 0.92$; Cohen's $d = 0.84$) or hit-KR ($M = 0.89$; Cohen's $d = 0.60$) achieved significantly higher sensitivity scores. The four KR conditions did not differ significantly from one another in perceptual sensitivity ($p > .05$ in each case).

Mean response bias scores for observers in the five KR conditions are plotted as a function of periods of watch in Fig. 3. The ANOVA indicated a significant effect for KR, $F(4, 105) = 2.54$, $p < .05$, $\omega^2 = .009$, and for period on watch, $F(3, 291) = 9.58$, $p < .001$, $\omega^2 = .06$. The interaction between these factors was not statistically significant ($p > .05$). Tukey HSD tests for the KR conditions revealed that observers receiving composite KR ($M = 0.91$) were significantly more conservative than those who did not receive KR ($M = 0.49$; Cohen's $d = 0.80$). All other comparisons failed to reach statistical significance ($p > .05$ in each case). The significant main effect for periods indicates that conservatism increased with time on watch (*M*s for Periods 1–6 were 0.54, 0.67, 0.71, 0.73, 0.77 and 0.77, respectively). These results are consistent with prior research in vigilance, and are probably due to growing awareness by participants during the vigil that the signals to be detected were relatively rare (Craig, 1978).

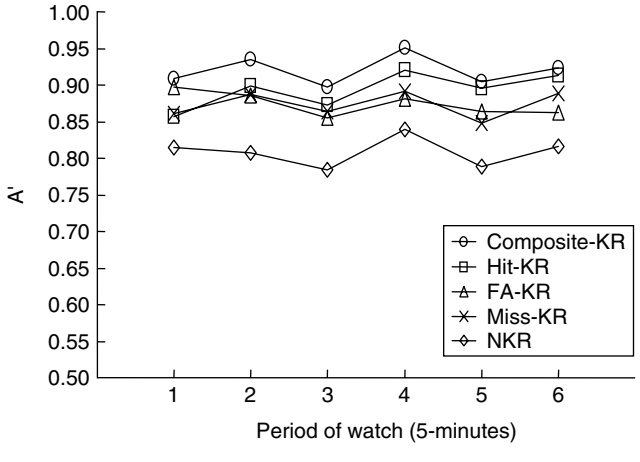


Figure 2. Perceptual sensitivity (A′) as a function of period on watch.
Note. NKR = no KR.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

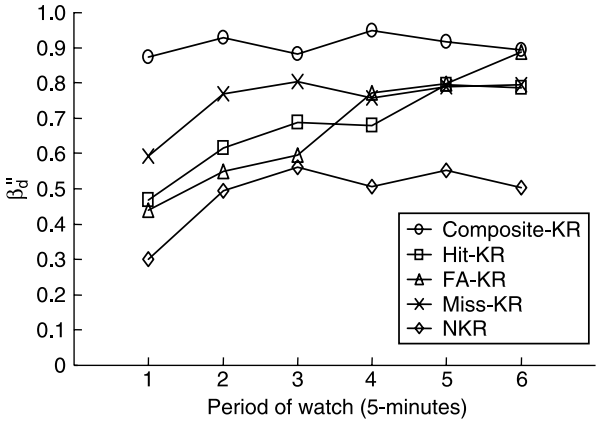


Figure 3. Response bias (β''_d) as a function of period on watch.
Note. NKR = no KR.

Perceived stress

Preliminary ANOVAs confirmed that participants across the KR conditions did not differ significantly in their pre-vigil *distress*, *task engagement* or *worry* scores ($p > .45$ in each case). Pre- and post-DSSQ responses were therefore used to compute pre-post z scores for the three DSSQ factors. One-sample t tests using the Bonferroni correction ($\alpha = .05$) indicated that across all conditions, observers reported significant increase in *distress* ($M = 0.66$), $t(109) = 6.42$, $p < .001$, and significant declines in *task engagement* ($M = -0.86$), $t(109) = -10.05$, $p < .001$, and *worry* ($M = -0.46$), $t(109) = -5.33$, $p < .001$. Thus, for two of the three dimensions of stress, the data indicate that the vigil itself was stressful. To examine KR effects, an ANOVA was computed for each stress scale. These analyses indicated that there were no significant differences among KR conditions for the *distress* and *worry* dimensions ($p > .05$ in each case), but that there was a significant KR effect for *task engagement*, $F(4, 105) = 3.59$, $p < .01$, $\omega^2 = .09$. Tukey HSD tests indicated that participants in the no-KR control condition ($M = -1.25$) reported a greater decline in *task engagement* after the vigil than those in the composite-KR condition ($M = -0.38$; Cohen's $d = 1.02$). No other comparisons reached conventional levels of significance ($p > .05$ in each case).

In sum, only hit- and composite-KR were effective in enhancing perceptual sensitivity relative to a no-KR control, and only composite-KR was effective in reducing the stress associated with the vigil. However, this effect was restricted to the *task engagement* dimension. In subsequent sections, the effects of pessimism and optimism are evaluated to determine whether they moderated these relations between KR, performance and stress state.

Pessimism and optimism

The mean pessimism score across all 110 participants was 36.5 ($SD = 6.6$), and the mean optimism score was 53.56 ($SD = 5.0$). The correlation between these variables was $r = -.40$ ($p < .001$), which is lower than that usually obtained between these two variables (approximately $r = .5$; see Dember, 2002; Dember *et al.*, 1989). However, inspection of the mean, standard deviation and coefficient alpha for each scale ($\alpha = .81$ for optimism; $\alpha = .83$ for pessimism) indicated that these indices did not meaningfully

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

differ from results obtained from prior samples. Preliminary analyses indicated that there were no significant differences among KR conditions in optimism or pessimism ($p > .5$ in each case).

Feedback and dispositional optimism and pessimism

To evaluate the impact of optimism and pessimism, and their interaction with KR format, a hierarchical regression approach was employed using procedures outlined by Pedhazur (1997; see also Hardy, 1993). The KR variable was dummy-coded so that the no-KR condition was assigned a value of zero across all four vectors. Each vector (i.e. hit-KR, false alarm-KR, miss-KR and composite-KR) therefore represented the comparison between a KR condition and the no-KR control. The terms in each regression were trait (optimism or pessimism), KR (the dummy coded vectors) and trait by KR product vector terms ('interaction' terms). Separate regressions were computed with optimism and pessimism to simplify analyses and avoid problems associated with multicollinearity.

When the vector products are included in the model, interpretation of the regression weights for KR and trait terms are meaningless, since the product vectors are correlated with their constituent parts (see Pedhazur, 1997; see also Jaccard & Turrisi, 2003). Thus, the interaction terms are interpreted according to the increment in the proportion of variance accounted for (i.e. ΔR^2) after the variance accounted for by pessimism and the KR vectors (the 'main effects') has been entered into the model. Step 1 of the hierarchical regression procedure consisted of entry of the trait variable (optimism or pessimism), Step 2 consisted of entry of the trait variable and the dummy coded KR variable, and Step 3 consisted of the main effects and the 'interaction' terms. A non-significant increment in ΔR^2 for the full model (Step 3) resulted in evaluating the models at Step 1 (trait only) or Step 2 (trait and KR). Cases in which a significant product vector was obtained indicated that the slope of the regression of the dependent variable on the trait variable was significantly different between the corresponding KR conditions. That is, the test for the regression weight for each product vector (e.g. hit-KR by pessimism, false alarm-KR by pessimism) constitutes a test of the difference between the regression weights for the no-KR group and the particular KR condition. Following Pedhazur, such product vectors were tested using the Johnson-Neyman procedure for determination of simultaneous regions of significance (see also Huitema, 1980; Jaccard & Turrisi, 2003). This procedure, analogous to a test of simple effects in ANOVA, involves determining over which regions of the continuous predictor variable (e.g. pessimism) differences between the categorical variables (e.g. no-KR vs. FA-KR) are significant. In this framework, the trait variable is considered a moderating variable for the relation between KR and the dependent variable (i.e. performance, stress; see Jaccard & Turrisi, 2003). To control for Type I error, the Bonferroni *F* statistic was used in the computation of regions of significance (see Huitema, 1980). To simplify performance analyses, regressions of performance on the trait variable (pessimism, optimism), KR and the associated product vectors were performed for sensitivity and response bias in two ways: (1) an overall index of sensitivity and bias based on the total number of correct detections and false alarms over the entire vigil; and (2) a difference score for these indices between the first and last periods on watch.

Pessimism

Performance

Regressions of overall A′ and the difference score between Periods 1 and 6 revealed no impact of pessimism on performance, nor did pessimism interact with KR in influencing

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

performance ($p > .05$ in each case). Similar analyses with overall β''_d and the difference score for Periods 1 and 6 indicated that pessimism did not impact response bias, nor did it interact with KR to influence overall response bias or change in bias between the beginning and end of the session ($p > .05$ in each case).

Perceived stress

Stress state was analysed via separate hierarchical regressions of the pre-post difference scores of each scale (*distress*, *task engagement* and *worry*) on pessimism, KR group and the interaction between these factors.

Distress

The regression of pre-post *distress* on pessimism, KR, and the product vectors (i.e. KR × pessimism) was significant, $F(9, 100) = 2.41$, $p < .05$, $R^2 = .18$. The increment in variance accounted for by the product vector terms was significant, $F(4, 100) = 2.59$, $p < .05$ $\Delta R^2 = .09$. Significant regression weights were observed for the dummy coded false alarm vector ($b = -5.35$, $SE_b = 2.06$, $\beta = -2.00$, $t(34) = -2.00$, $p < .05$) and the product vector for pessimism and false alarm-KR ($b = 0.14$, $SE_b = 0.06$, $\beta = 1.92$, $t(34) = 2.46$, $p < .05$). The significant effect for the regression weight associated with the product vector for pessimism and false alarm-KR indicates that the regression weights for the no-KR and false alarm-KR conditions were significantly different from one another. The relation between pre-post task *distress* and pessimism is shown in Fig. 4 for the false alarm-KR and no-KR conditions. Application of the Johnson-Neyman procedure revealed that for pessimism scores less than 31.7, observers who received false alarm-KR reported significantly less *distress* than those in the no-KR group. Hence,

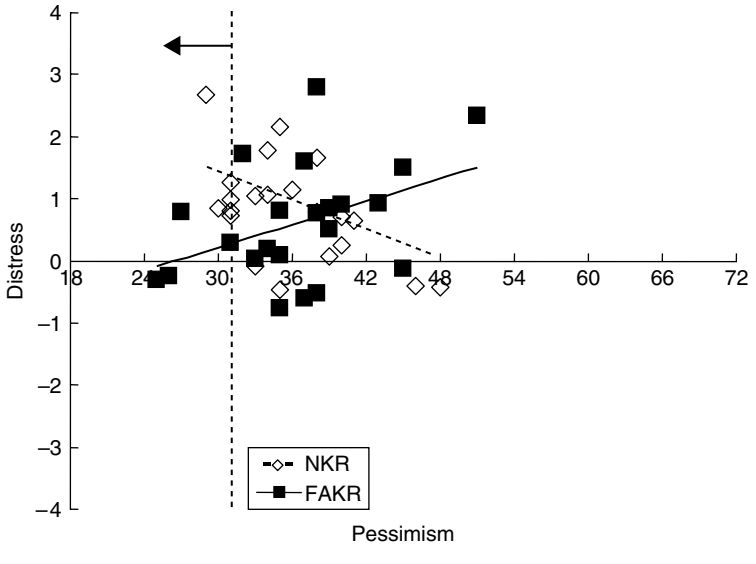


Figure 4. Pre-post task *distress* as a function of pessimism for the false alarm-KR and no-KR groups. *Note.* Dotted line with arrow represents the region of significant differences between the regression lines. Separate regression equations: false alarm-KR group: D = .06P − 1.64, $R^2 = .17$; no-KR group: D = −.08P + 3.71, $R^2 = .13$; intersection point of regression lines: P = 38.8. NKR = no-KR; FAKR = false alarm-KR; D = distress; P = pessimism.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

false alarm-KR was effective in reducing *distress*, but only for participants low in pessimism. This effect was attenuated for individuals high in pessimism.

Task engagement

The regression of pre-post *task engagement* on pessimism, KR and the product vectors was significant, $F(9, 100) = 3.57$, $p < .01$, $R^2 = .24$. The increment in variance accounted for by the product vector terms was significant, $F(4, 100) = 3.33$, $p < .05$ $\Delta R^2 = .11$. Significant regression weights were observed for the dummy coded false alarm vector ($b = 4.64$, $SE_b = 1.64$, $\beta = 2.08$, $t(34) = 2.82$, $p < .01$) and the product vector for pessimism and false alarm KR ($b = -0.11$, $SE_b = 0.04$, $\beta = -1.85$, $t(34) = -2.47$, $p < .05$). Pre-post *task engagement z* scores are plotted as a function of pessimism for the false alarm-KR and no-KR conditions in Fig. 5. The Johnson-Neyman procedure was applied to investigate the significant false alarm-KR/pessimism product vector. This analysis revealed that for pessimism scores below 35.6, observers who received false alarm-KR reported significantly higher *task engagement* than observers who did not receive feedback. That is, false alarm-KR was more effective in increasing *task engagement* for participants relatively low in pessimism, but the effectiveness of this KR format was reduced for individuals higher in pessimism.

Worry

Analyses indicated non-significant regressions for pessimism and KR ($p > .05$ in each case), but a significant regression for the full model (Step 3), $F(9, 100) = 2.23$, $p < .05$, $R^2 = .18$. The increment in variance accounted for by the product vector terms was significant, $F(4, 100) = 3.79$, $p < .01$ $\Delta R^2 = .13$. This outcome is analogous to non-significant main effects in an ANOVA but a significant interaction. However, tests of the

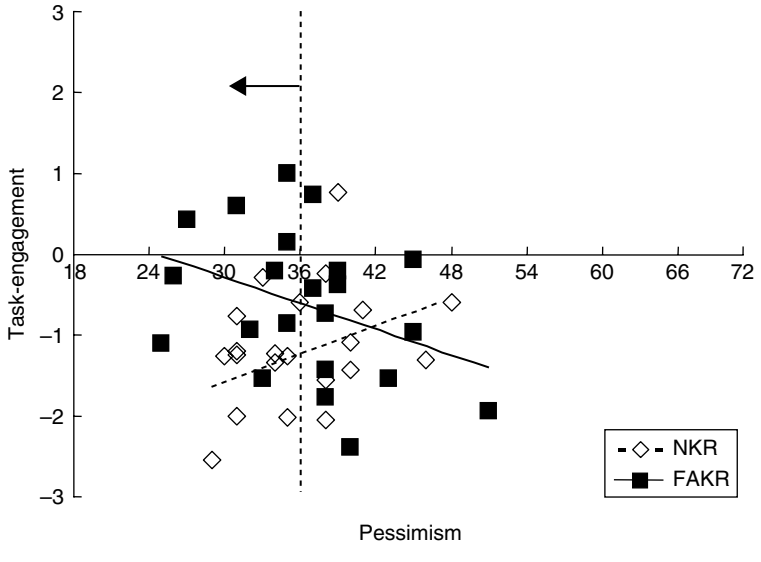


Figure 5. Pre-post *task engagement* as a function of pessimism for the false alarm-KR and no-KR groups. *Note.* Dotted line with arrow represents the region of significant differences between the regression lines. Separate regression equations: false alarm-KR group: TE = −.05P + 1.32, $R^2 = .14$; no-KR group: TE = .06P − 3.32, $R^2 = .11$; intersection point of regression lines: P = 41.8. NKR = no-KR; FAKR = false alarm-KR; TE = task engagement; P = pessimism.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

regression weights revealed no significant *b*s, although the coefficient for the product vector for Miss-KR/pessimism was marginally significant ($b = -0.08$, $SE_b = 0.05$, $\beta = -1.41$, $t(34) = -1.87$, $p = .06$). Pre-post *worry* scores are plotted as a function of pessimism for the miss-KR and no-KR conditions in Fig. 6. Application of the Johnson-Neyman procedure was unsuccessful, probably due to high within-groups variability, which can cause failures in solutions for regions of significance (see Pedhazur, 1997). Separate regressions of *worry* on pessimism for the no-KR and miss-KR conditions revealed a non-significant regression for the former and a significant regression for the latter, $F(1, 20) = 9.11$, $p < .01$, $R^2 = .31$ ($b = -0.06$, $SE_b = 0.02$, $\beta = -56$, $t(34) = -3.09$, $p < .01$). Contrary to prediction, increased pessimism was related to decreased *worry* for participants in the miss-KR condition.

Optimism

Performance

Regressions of overall A′ on optimism, KR and their product vectors revealed a significant regression for the full model, $F(9, 100) = 2.73, p < .01$. However, inspection of the data indicated that these effects were primarily due to two outlier scores in the false alarm-KR condition. When these were omitted from the analysis, no significant effects were observed. Inspection of the DSSQ and OPI scores of these two individuals indicated that they did not differ from their group mean (the false alarm-KR condition) on the OPI or DSSQ scales by more than 1.5 standard deviations. Similar analyses on A′ difference scores, β''_d and β''_d difference scores indicated that optimism did not significantly influence change in performance between the beginning and end of the vigil, nor did optimism interact with KR format in influencing overall response bias or changes or response bias ($p > .05$ in each case).

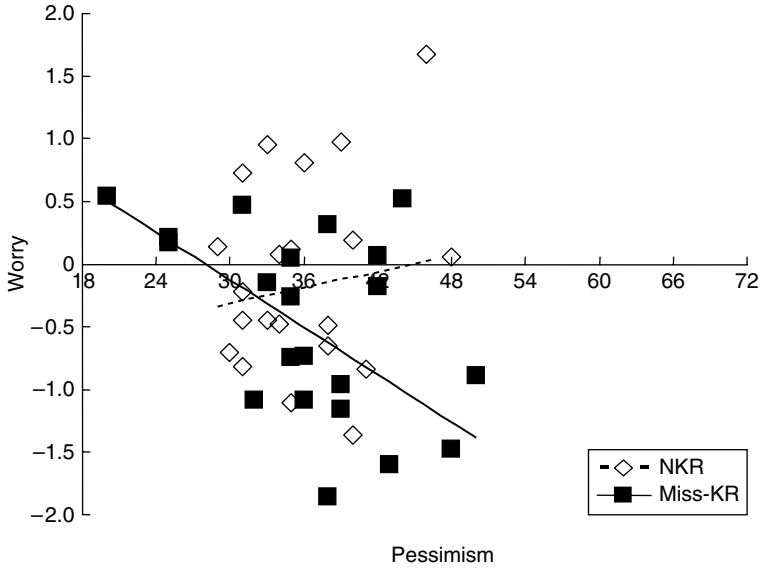


Figure 6. Pre-post task *worry* as a function of pessimism for the miss-KR and no-KR groups. *Note.* Separate regression equations: miss-KR group: $W = -.06P + 1.76$, $R^2 = .31$; no-KR group: $W = .02P - .96$, $R^2 = .01$; intersection point of regression lines: $P = 32.3$. NKR = no-KR; W = worry; P = pessimism.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Perceived stress

As with the analyses involving pessimism, stress state was analysed via separate hierarchical regressions of each scale (*distress*, *task engagement* and *worry*) on optimism, KR and the interaction between these factors.

Distress and task engagement

Regression analysis indicated that optimism, KR, and the product vectors did not significantly predict *distress* ($p > .05$). Optimism predicted pre-post *task engagement*, $F(1, 108) = 6.50$, $p < .05$, $R^2 = .06$, and the model including optimism and KR conditions was significant, $F(5, 104) = 3.84$, $p < .05$, $R^2 = .16$, although the product terms for these variables did not significantly add to the variance accounted for by the regression ($p > .05$). Significant regression weights were observed for optimism, ($b = -0.03$, $SE_b = 0.02$, $\beta = -0.19$, $t(38) = -2.08$, $p < .05$), the false alarm-KR vector ($b = 0.56$, $SE_b = 0.26$, $\beta = 0.25$, $t(38) = 2.18$, $p < .05$) and the composite-KR vector ($b = 0.79$, $SE_b = 0.26$, $\beta = 0.36$, $t(38) = 3.09$, $p < .01$).

Worry

For pre-post task *worry*, the full regression model was significant, $F(9, 100) = 2.62$, $p < .01$, $R^2 = .19$, and the product vectors significantly added to the variance accounted for by the regression, $F(4, 100) = 4.83$, $p < .01$, $\Delta R^2 = .16$. Significant regression weights were observed for the false alarm-KR vector ($b = 7.39$, $SE_b = 2.72$, $\beta = 3.26$, $t(34) = 2.72$, $p < .01$), and the false alarm-KR/optimism product vector ($b = -0.14$, $SE_b = 0.05$, $\beta = -3.34$, $t(34) = -2.84$, $p < .01$). Pre-post *worry* scores associated with the false alarm-KR and no-KR conditions are plotted as a function of optimism in Fig. 7.

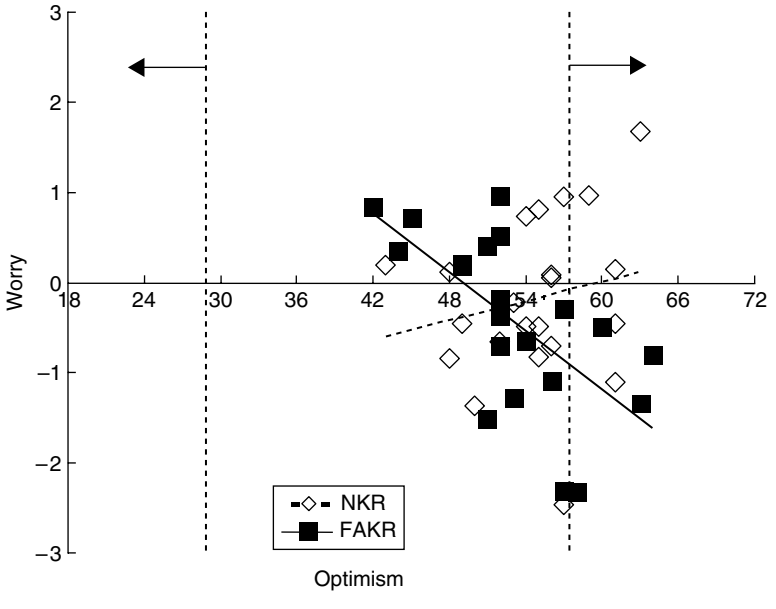


Figure 7. Pre-post task *worry* as a function of optimism for the false alarm-KR and no-KR groups. *Note*. Dotted lines with arrows represent the simultaneous regions of significant differences between the regression lines. Separate regression equations: false alarm-KR group: $W = -.11Opt + 5.26$, $R^2 = .41$; no-KR group: $W = .04Opt - 2.12$, $R^2 = .04$; intersection point of regression lines: Opt = 51.8. NKR = no-KR; FAKR = false alarm-KR; W = worry; Opt = optimism.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Johnson-Neyman analyses revealed that for optimism scores below 28.6, false alarm-KR induced greater *worry* compared with No-KR, but for scores above 57.6, receiving false alarm-KR induced less worry relative to no-KR controls. Note that the results for the low end of the optimism scale must be interpreted with caution since the regions of significance represent an extrapolation to a region in which there are no sample data.

Discussion

A primary purpose of the present study was to investigate whether the performance differences as a function of KR format observed in previous research (Dittmar *et al.*, 1985) were associated with differences in stress state, and whether combining these formats into a single condition is most effective in improving performance and reducing stress. A second goal was to investigate the possibility, based upon hypotheses derived from FIT (Kluger & DeNisi, 1996), that dispositional optimism and pessimism would influence the effectiveness of different KR formats in improving vigilance performance and in reducing the stress associated with prolonged monitoring. Specifically, it was hypothesized that false alarm- and miss-KR should impair performance and increase stress for individuals high in pessimism. By contrast, hit-KR should facilitate performance and reduce stress symptoms for individuals high in optimism, although these effects were expected to be weaker than those associated with pessimism.

The present results obtained with hit-KR were consistent with the findings of Dittmar *et al.* (1985), while the composite-KR was effective at enhancing sensitivity as also observed in prior testing (Szalma *et al.*, 1998, 2004). Also consistent with Dittmar *et al.,* miss-KR was not effective in enhancing perceptual sensitivity. However, in contrast to the results of Dittmar and her colleagues, false alarm-KR was also not effective in enhancing sensitivity relative to a no-KR control. Moreover, a major finding of Dittmar and her colleagues of an interaction between KR format and time on task for both sensitivity and response bias was not confirmed in the present experiment. This is most probably due to the absence here of a sensitivity decrement and large within-groups variability observed in the response bias data. These differences are probably task-specific and confirm that the influence of KR format depends on the information-processing requirements of the particular vigilance task at hand. The task employed by Dittmar *et al.* required discrimination of the length of a vertical line, while in this experiment, participants monitored patterns of change in the shape of an object display. Prior research has shown that use of an object display enhances sensitivity and attenuates the vigilance decrement without the provision of feedback (Szalma, 2002b). It might be that the differential effects of KR format depend on the perceptual and cognitive demands of the specific task, and while several of these factors have been determined (Davies & Parasuraman, 1982; Warm & Dember, 1998), a full exposition has yet to emerge. In comparison with the Dittmar *et al.* results, note that the mode of KR also has varied since their report. Dittmar and her colleagues used a 1,000 Hz tone as feedback, while in the present experiment a female voice controlled by the computer was employed to deliver feedback. Whether such mode differences in KR presentation influences effectiveness for enhancing vigilance is a matter for future research.

Stress

Consistent with previous vigilance experiments, observers in this study found the vigil to be stressful (Hancock & Warm, 1989). A significant increase in *distress* and reduction

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

in *task engagement* was observed across all conditions. However, KR format failed to significantly impact *distress* or *worry*. A significant KR effect was observed for *task engagement*, but *post hoc* tests revealed that this effect was restricted to the composite-KR condition. Taken together with the performance results, these results indicate that although hit-, miss-, and false alarm-KR differ in their effects, combining them has the greatest impact in improving performance and reducing stress. Moreover, these results indicate that regardless of the separate effects of the different KR formats, training regimens for vigilance may benefit from providing feedback regarding all three outcomes. Providing such information may direct attention away from the self and towards task goals to a greater degree than when the formats are presented singly. The different forms of KR in this instance might serve as cues for observers to attend to their respective responses to stimuli, so that the KR serves to strengthen the association between stimulus characteristics and observer response. The different effects across the three stress state dimensions reflect the multidimensional nature of the stress construct (Matthews, 2001), and may represent different appraisals to task demands and different coping responses to those demands. Thus, the 'core relational theme' (Lazarus, 1991) that links to *distress* is 'overload of capacity for successful performance' and *task engagement* relates to 'commitment of effort' (Matthews *et al.*, 2002, 1999). Thus, composite-KR did not relieve the overload associated with vigilance (see Warm *et al.*, 1996), but it did enhance the commitment of effort to perform. The latter result conforms to the exposition concerning the motivational effect that KR has on vigilance (see Warm & Jerison, 1984).

In this experiment, *worry* declined after the vigil across all conditions. According to Matthews *et al.* (1999) the *worry* dimension is associated with self evaluation 'such that the nature of the task provokes re-assessment of personal qualities and goals' (p. 348) and with emotion-focus and avoidant coping (Matthews & Campbell, 1998). Indeed, they noted that self-reflection is most likely to occur in monotonous tasks such as vigilance. Further, Matthews *et al.* (2002) argued that this theme would fall under the *meta-task* process described by Kluger and DeNisi (1996). In this experiment, *worry* declined across all conditions, indicating an overall attenuation of *meta-task* processes. However, these effects depended on the pessimism and optimism of observers, as discussed below.

Performance and stress: The non-reciprocity of hit- and false alarm-KR

Dittmar and her colleagues noted that hit- and false alarm-KR are reciprocals of one another since they provide the same information regarding whether a response was correct. However, their effects in this study were not equivalent in terms of performance and self-report of stress state. Hit-KR, but not false alarm-KR, enhanced sensitivity relative to observers who did not receive KR. False alarm-KR, but not hit-KR, reduced the *distress* and increased *task engagement* relative to no-KR controls, but only for individuals low in pessimism. Although hit-KR observers performed better than controls who did not receive KR, this performance enhancement was not associated with reduced *distress* or increased *task engagement*, suggesting that hit-KR does not alter the coping strategies observers use to deal with task demands. In contrast, false alarm-KR reduced the stress symptoms for individuals low in pessimism, but this benefit was not associated with performance effects. This indicates that false alarm feedback influences appraisals observers make in respond to task demands, that these appraisals occur independent of actual performance, and that these appraisals depend upon the person's level of dispositional pessimism.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Effects of pessimism and optimism: The test of feedback intervention theory (FIT)

Performance

Consistent with prior experimental findings, pessimism and optimism did not significantly affect the performance of observers (see Ganey *et al.*, 2003; Szalma, 2002a; Thropp, Szalma, Ross, & Hancock, 2003). In one exception, Helton *et al.* (1999) observed a steeper decrement in pessimists relative to optimists, but evidence from this and other investigations implies that these effects may be dependent on task characteristics. Specifically, Helton *et al.* employed a rather difficult discrimination at a very high event rate (57.5 per minute), while relatively easier discriminations at lower event rates were used in this and other experimental procedures which did not show this influence.

Helton and his colleagues also observed that pessimism/optimism did not interact with false summary KR to influence performance, a finding that was replicated in this study using veridical feedback after each response (or failure to respond appropriately). This result was inconsistent with the hypotheses based on FIT. One possible explanation is that the nature of the KR and the task was not sufficiently important relative to the individuals' self-goals (i.e. *meta-task* goals). This interpretation is supported by the finding that across all conditions, pre-post *worry* declined, which reflects less self-evaluative processes (Matthews *et al.*, 2002). Stronger effects might be observed if the attention to the levels of the goal hierarchy discussed by Kluger and DeNisi (1996; i.e. *meta-task*, *task-motivational* and *task-learning processes*) were manipulated specifically.

A second explanation can be derived from the processing efficiency theory of Eysenck and Calvo (1992). According to this theory, individuals who have difficulty coping with stress can compensate by exerting more task effort. The effects of stress are therefore manifested as a reduction in the efficiency of information processing rather than in the output of such processing. Thus, two individuals may exhibit equivalent levels of performance, but one individual is maintaining that performance at greater cost. Evidence from the current study indicates that individuals high in pessimism can achieve levels of performance equivalent to those of their less pessimistic cohorts but that maintaining such performance occurs at greater cost, particularly when negative feedback is provided. It may therefore be that negative KR and pessimism interact to influence processing efficiency rather than the absolute level of performance.

Stress

Pessimism

Although the expected interaction between trait and KR was not observed with regard to performance, interactions were observed in stress state and the effect of pessimism on stress depended on the joint effects of KR format and dimension of stress. Thus, the effect of 'negative KR' (i.e. false alarm- and miss-KR) in reducing stress was observed only for individuals low in pessimism. Moreover, false alarm-KR exerted its influence on *distress* and *task engagement*, while miss-KR exerted its influence on *worry*. However, while false alarm-KR reduced *distress* and increased *task engagement* for individuals low in pessimism, miss-KR reduced *worry* for individuals *higher* in that trait. Thus, the two forms of negative KR influenced different stress state dimensions in different ways. It is important to remember that *distress* and *task engagement* dimensions relate to mood, motivation and cognitive states, but the *worry* dimension reflects only the cognitive component of affective states (Matthews *et al.*, 2002, 1999). False alarm-KR

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

may therefore have pervasive effects across the mood, motivation and cognitive components of state, while miss-KR may influence primarily the cognitive component. The latter result does not conform to the original hypotheses, based on the findings of Dittmar *et al.* (1985) and Kluger and DeNisi (1996) that one would expect miss-KR would cause participants to engage in more of the *meta-task* processes associated with the *worry* dimension (Matthews *et al.*, 2002), and that these effects would be more likely for those higher in pessimism. However, individuals higher in pessimism who received miss-KR reported *less worry* after the vigil, indicating that they were engaging in less self-evaluative processes.

The interaction of KR with pessimism may also be influenced by the different coping strategies associated with the three dimensions of stress state. *Distress* and the absence of *task engagement* are associated with emotion-focused coping (Matthews & Campbell, 1998; Matthews *et al.*, 1999), and this form of coping is also typically adopted by pessimists (Helton *et al.*, 1999; Scheier & Carver, 1987). Further, Szalma (2002a) indicated that pessimism is associated with increased reliance on emotion-focused coping strategies. Thus, the differences between false alarm-KR and no-KR on the *distress* and *task engagement* dimensions may be due to less emotion focused coping among those low in pessimism. As pessimism increases, emotion focused coping increases, diverting processing resources to *meta-task processes* (Kluger & DeNisi, 1996), and thereby reduces the effectiveness of KR in reducing these symptoms.

Optimism

As expected, the interaction effects between KR and optimism were weaker than those between KR and pessimism. For participants who received false alarm-KR, higher levels of optimism were associated with decreased *worry* and lower levels of optimism associated with increased *worry*. However, higher optimism predicted lower *task engagement* across all KR conditions. Taken together, these results indicate that individuals high in optimism find false alarm-KR less aversive than those low on that trait, but optimism does not protect against loss of engagement. The reasons for the latter effect are a matter for future research.

Summary and conclusions

It is evident that the stress effects of vigilance are complex and multidimensional and that the effectiveness of KR in reducing stress depends on both the type of KR provided and characteristics of the individual. Taken together with the results of prior research in this area, the current study indicates that these KR effects depend on the perceptual demands of the task. The results of the present study also serve to underscore the importance of examining individual differences in stress responses to vigilance (Berch & Kanter, 1984). KR format did not significantly impact *distress* or *worry* dimensions of stress, but KR effects on these variables did emerge when dispositional pessimism and optimism were taken into account. Further investigation needs to be directed to explore these joint effects of feedback and observer characteristics on performance and stress in vigilance tasks. Such efforts can lead to improved training regimens that are tailored to the individual to maximize performance in crucial operational settings.

The hypotheses based upon FIT received partial support in that negative feedback did not enhance performance, and that false alarm-KR was effective in reducing stress symptoms only for those individuals low in pessimism. However, the predicted

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

interaction effect on performance of feedback type and pessimism was not observed. Future investigations should test for this interaction by systematically manipulating participants' goals according to the hierarchy described in Kluger and DeNisi (1996).

Since 9/11, the world has become much more wary and invested in safety and security systems. Often, these individuals and the technology created to help them perform their tasks are looking for the equivalent of a 'needle in a haystack'. Even with highly sensitive systems and highly motivated observers, such rare signal occurrences lead to an abundance of false alarms (Parasuraman, Hancock, & Olofinboba, 1997) that eventually society in general will not tolerate. From our present experiment, it is clear that the effectiveness and efficiency of vigilance is contingent upon the character of the targets being searched, how that information is presented, and the nature of the individual to whom it is presented. The current findings point to the strong, interactive nature of these factors and begin to chart a course to better systems, better selection and better training for those who perform this vital task.

Acknowledgements

Completion of this work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-01-1-0621. P. A. Hancock, Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect official army policy. The authors wish to thank Dr Sherry Tove, Dr Elmar Schmeisser and Dr Mike Drillings for providing administrative and technical direction for the grant.

References

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors*, *34*, 513–533.

Bennett, K. B., Nagy, A. L., & Flach, J. M. (1997). Visual displays. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 659–696). New York: Wiley.

Berch, D. B., & Kanter, D. R. (1984). Individual differences. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 143–178). Chichester: Wiley.

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control theory approach to human behavior*. New York: Springer.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. New York: Cambridge University Press.

Chang, E. C. (Ed.). (2002). *Optimism and pessimism: Implications for theory, research, and practice*. Washington DC: American Psychological Association.

Craig, A. (1977). Broadbent and Gregory revisited: Vigilance and statistical decision. *Human Factors*, *19*, 25–36.

Craig, A. (1978). Is the vigilance decrement simply a response toward probability matching? *Human Factors*, *20*, 441–446.

Craig, A. (1979). Nonparametric measures of sensory efficiency for sustained monitoring tasks. *Human Factors*, *21*, 69–78.

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Dember, W. N. (2002). The optimism-pessimism instrument: Personal and social correlates. In E. C. Chang (Ed.), *Optimism and pessimism: Implications for theory, research, and practice* (pp. 281–299). Washington, DC: American Psychological Association.

Dember, W. N., Martin, S. H., Hummer, M. K., Howe, S. R., & Melton, R. S. (1989). The measurement of optimism and pessimism. *Current Psychology: Research and Reviews*, *8*, 102–119.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Dittmar, M. L., Warm, J. S., & Dember, W. N. (1985). Effects of knowledge of results on performance in successive and simultaneous vigilance tasks: A signal detection analysis. In R. E. Eberts & C. G. Eberts (Eds.), *Trends in ergonomics/human factors II* (pp. 195–202). Amsterdam: Elsevier.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*, 275–277.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion*, *6*, 409–434.

Ganey, H. C. N., Szalma, J. L., Hancock, P. A., Mouloua, M., Connolly, A., Dalton, J., & Davis, F. (2003). The role of choice and individual differences in the mitigation of noise and task stressors in vigilance. *Proceedings of the Human Factors and Ergonomics Society*, *47*, 1078–1081.

Hancock, P. A., & Hart, S. G. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in Design*, *10*, 6–16.

Hancock, P. A., & Szalma, J. L. (2003). Operator stress and display design. *Ergonomics in Design*, *11*, 13–18.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*, 519–537.

Hardy, M. A. (1993). *Regression with dummy variables*. Newbury Park, CA: Sage.

Harris, D. (2002). How to really improve airport security. *Ergonomics in Design*, *10*, 17–22.

Helton, W. S., Dember, W. N., Warm, J. S., & Matthews, G. (1999). Optimism-pessimism and false failure feedback: Effects on vigilance performance. *Current Psychology*, *18*, 311–325.

Huitema, B. E. (1980). *The analysis of covariance and alternatives*. New York: Wiley.

Hummer, M. K., Dember, W. N., Melton, R. S., & Schefft, B. K. (1992). On the partial independence of optimism and pessimism. *Current Psychology: Research and Reviews*, *11*, 37–50.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed). Thousand Oaks, CA: Sage.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Matthews, G. (2001). Levels of transaction: A cognitive science framework for operator stress. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 5–33). Mahwah, NJ: Erlbaum.

Matthews, G., & Campbell, S. E. (1998). Task-induced stress and individual differences in coping. *Proceedings of the Human Factors and Ergonomics Society*, *42*, 821–825.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., Grier, R., & Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, *2*, 315–340.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state big three? In I. Mervielde, I. J. Deary, F. DeFruyt & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 335–350). Tilburg: Tilburg University Press.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Erlbaum.

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision warning systems. *Ergonomics*, *39*, 390–399.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed). Ft. Worth: Harcourt Brace.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Scheier, M. F., & Carver, C. S. (1987). Dispositional optimism and physical well being: The influence of generalized outcome expectancies on health. *Journal of Personality*, *55*, 170–210.

See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). A meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin*, *117*, 230–249.

See, J. E., Warm, J. S., Dember, W. N., & Howe, S. R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors*, *39*, 14–29.

Szalma, J. L. (2002a). Individual differences in the stress and workload of sustained attention. *Proceedings of the Human Factors and Ergonomics Society*, *46*, 1002–1006.

Szalma, J. L. (2002b). Workload and stress of configural displays in vigilance tasks. *Proceedings of the Human Factors and Ergonomics Society*, *46*, 1536–1540.

Szalma, J. L., Hancock, P. A., Warm, J. S., Dember, W. N., & Parsons, K. S. (in press). Training for vigilance: Using predictive power to evaluate feedback effectiveness. *Human Factors.*

Szalma, J. L., Miller, L. C., Hitchcock, E. M., Warm, J. S., & Dember, W. N. (1998). Intraclass and interclass transfer of training for vigilance. In M. W. Scerbo & M. Mouloua (Eds.), *Automation Technology and Human Performance* (pp. 183–187). Mahwah, NJ: Erlbaum.

Szalma, J. L., Warm, J. S., Matthews, G., Dember, W. N., Weiler, E. M., Meier, A., & Eggemeier, F. T. (2004). Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Human Factors*, *46*, 219–233.

Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors*, *42*, 183–194.

Thropp, J. E., Szalma, J. L., Ross, J. M., & Hancock, P. A. (2003). Individual differences in dispositional pessimism, stress, and coping as a function of task type. *Proceedings of the Human Factors and Ergonomics Society*, *47*, 1073–1077.

Warm, J. S. (1993). Vigilance and target detection. In B. M. Huey & C. D. Wickens (Eds.), *Workload transition: Implications for individual and team performance* (pp. 139–170). Washington DC: National Academy Press.

Warm, J. S., & Dember, W. N. (1998). Tests of vigilance taxonomy. In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 87–112). Washington DC: American Psychological Association.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance* (pp. 183–200). Mahwah, NJ: Erlbaum.

Warm, J. S., Epps, B. D., & Ferguson, R. P. (1974). Effects of knowledge of results and signal regularity on vigilence performance. *Bulletin of the Psychonomic Society*, *4*, 272–274.

Warm, J. S., & Jerison, H. J. (1984). The psychophysics of vigilance. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 15–59). Chichester, UK: Wiley.

Warm, J. S., Kanfer, F. H., Kuwada, S., & Clark, J. L. (1972). Motivation in vigilance: Effects of self-evaluation and experimenter controlled feedback. *Journal of Experimental Psychology*, *92*, 123–127.

Received 20 April 2005; revised version received 14 July 2005

Cognitive Changes in the Infantry Field Environment

W.C. Harris & P.A. Hancock



Focus

- Military personnel
- Multiple, military relevant stressors
- Extended exposure to high stress level
- Measurement in situ



Field Studies

- Ft. Chaffee 2002
- Ft. Chaffee 2003
- Ft Bragg Army SERE School



Ft. Chaffee 2002

- Week 1 of field exercises
  - No accuracy or response time change
  - Trends
    - Deficits - Alpha greater than Charlie
    - Spatial Processing -faster with more errors
- Post-Assault
  - Increased SRT
- Medical incident
  - Increased SRT
  - Faster, less accurate information processing

Ft. Chaffee 2003

- Week 1 of Field Exercises
  - SRT – Increased and remained slower
  - Information Processing
    - Initially slow and accurate
    - Returned to baseline speed but accuracy declined
    - Not all tasks affected equally
- Post-Assault
  - SRT increased
  - LRS & Math slower
  - No accuracy change

Ft Bragg Army SERE School

W. C. Harris & P.A. Hancock

University of Central Florida

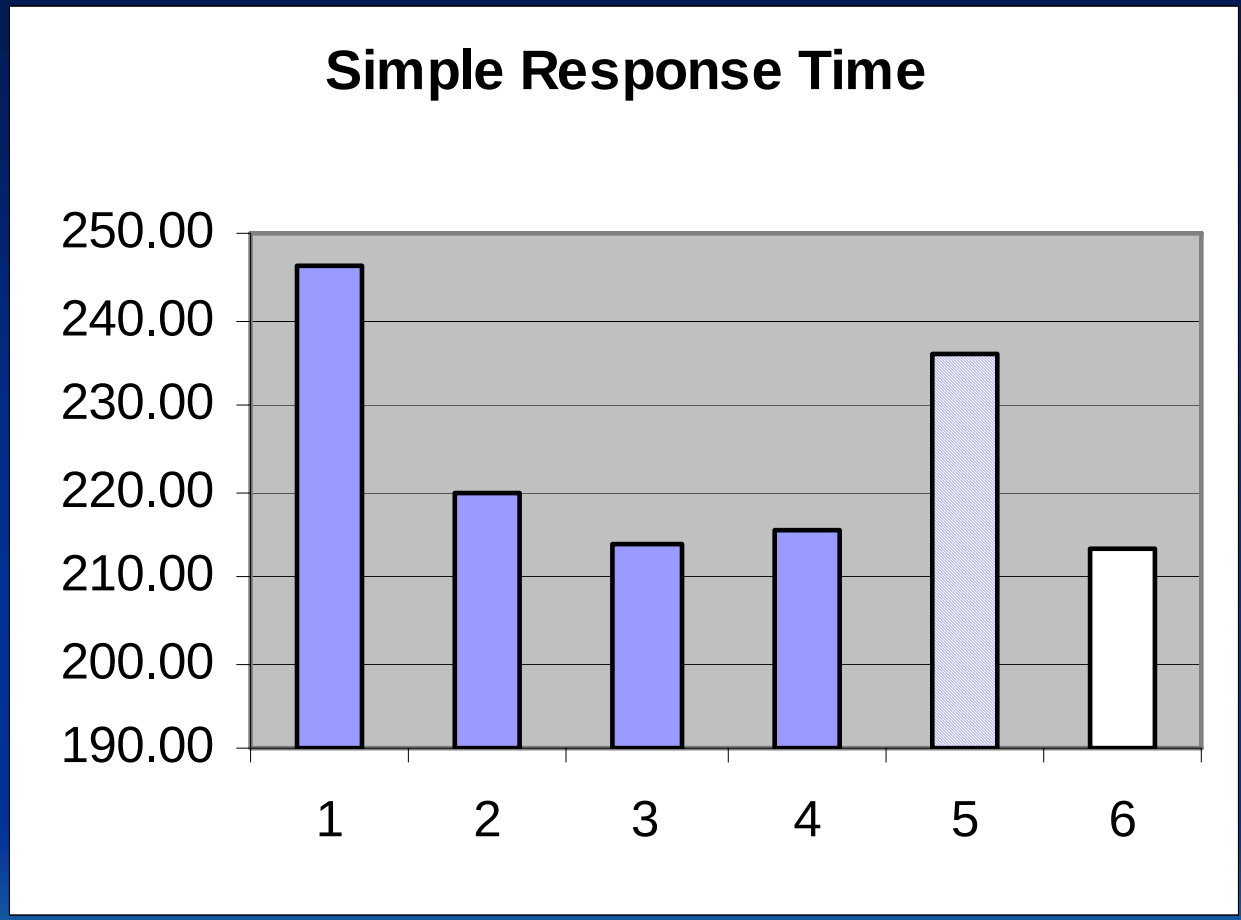
C.A. Morgan III

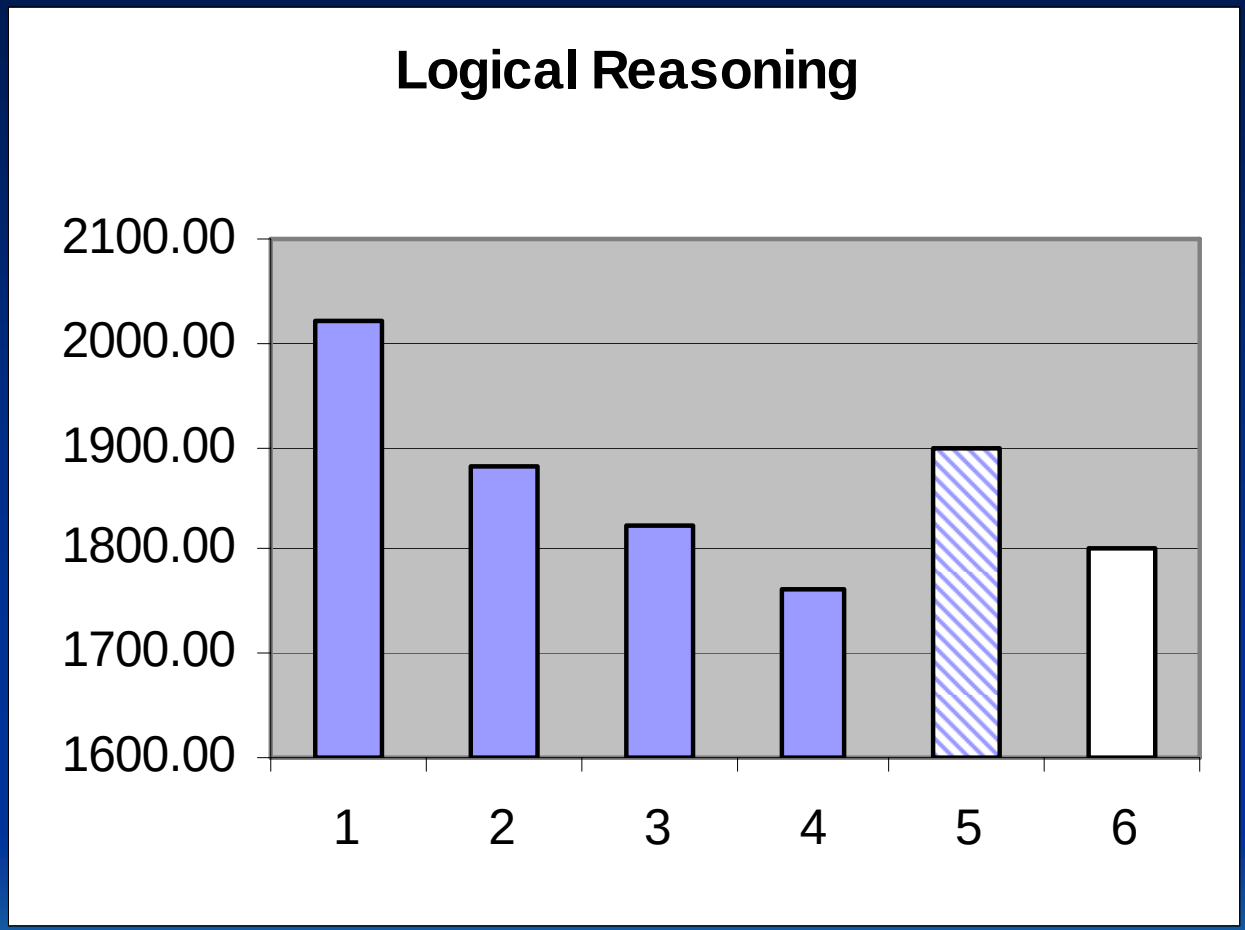
Yale University School of Medicine

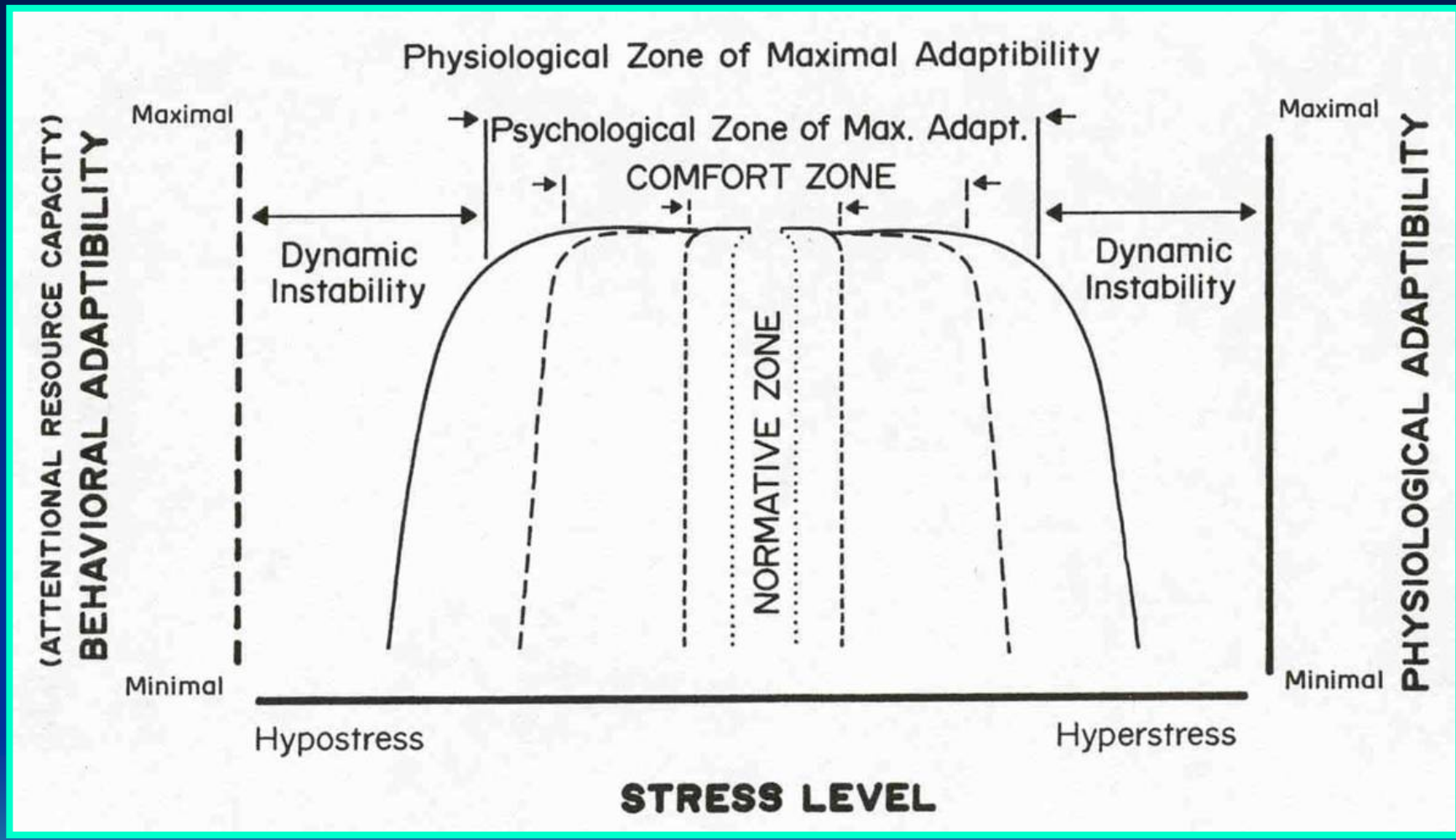


- Four pre-stress assessments
- Acute effect
  - SRT increased
  - Slower Information-Processing
  - Stable accuracy except mathematical processing (decreased)
- Complete recovery with 12 hours of rest




Simple Response Time
250.00
240.00
230.00
220.00
210.00
200.00
190.00
1
2
3
4
5
6


Logical Reasoning
2100.00
2000.00
1900.00
1800.00
1700.00
1600.00
1
2
3
4
5
6



Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31(5), 519-537.

Conclusions

- Performance remains stable in Infantry Operating Conditions
- As stress become greater (SERE School) accuracy is maintained but speed decreases
- Fast, inaccurate performance appears during extreme stress
- Performance recovers rapidly following stress
- Simple Response Time is a sensitive measure of fatigue
- Motivation appears to minimize the effects of stress

Cognitive Performance During 30 Day Live-Fire Exercise

W.C. Harris & P.A. Hancock

University of Central Florida

Steve Murray & Wayne Ensign

SPAWAR, San Diego

Procedures

- Participants
- Environmental Conditions
- Cognitive Assessment Schedule
- Compliance
- Cognitive Assessment Battery



ARES-I Scales

- Respond to all silhouettes (SRT)
- Only respond to target silhouettes
- Classify all silhouettes (2 Choice)
- Logical Reasoning - Convoy
- Map - Memory & Rotation
- Match to Sample
- Mood Scale

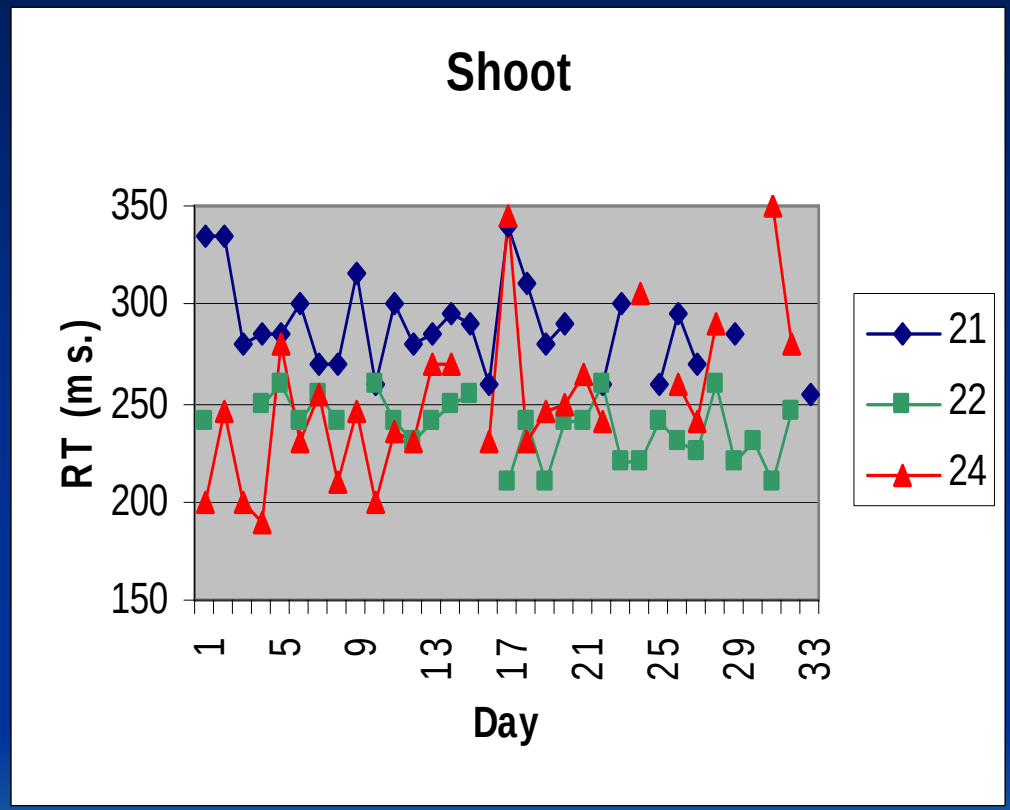


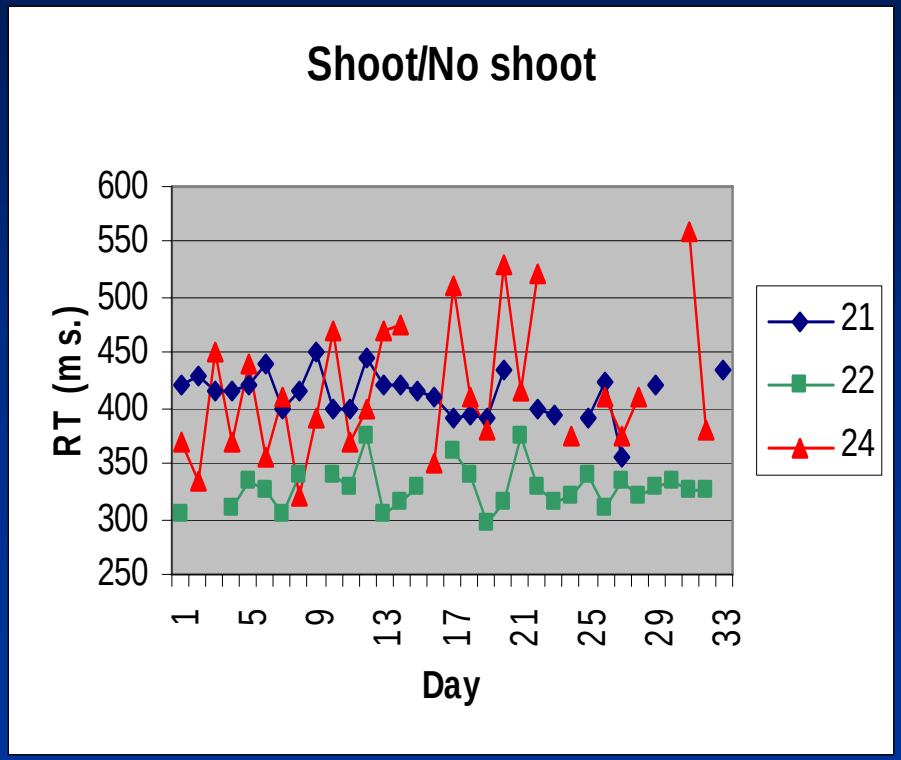
Results

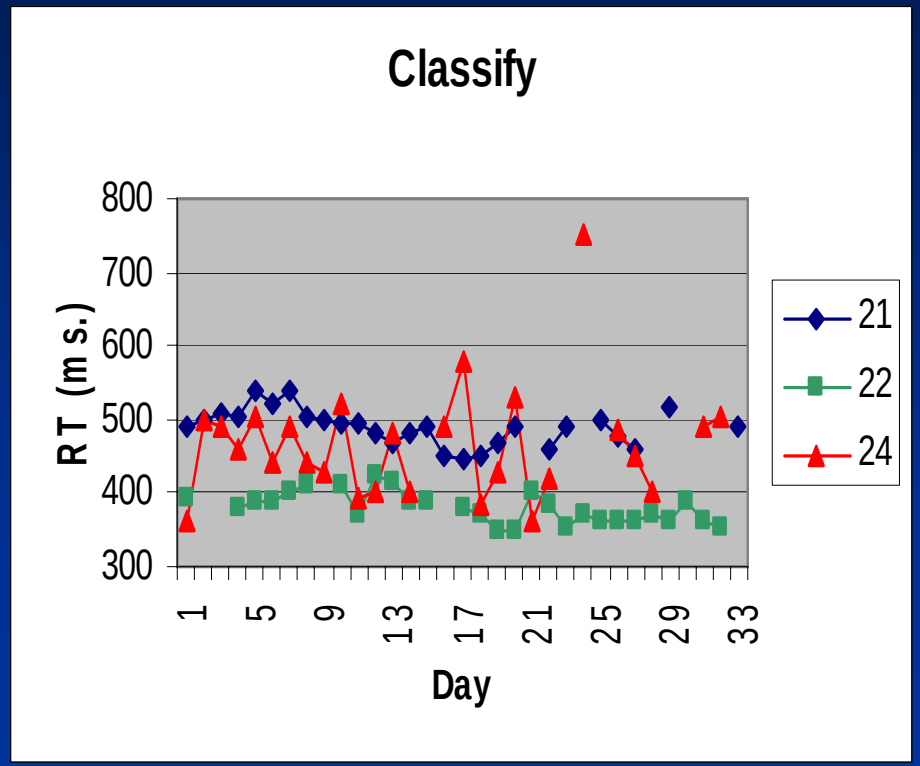
- Response Time
- Accuracy
- Mood

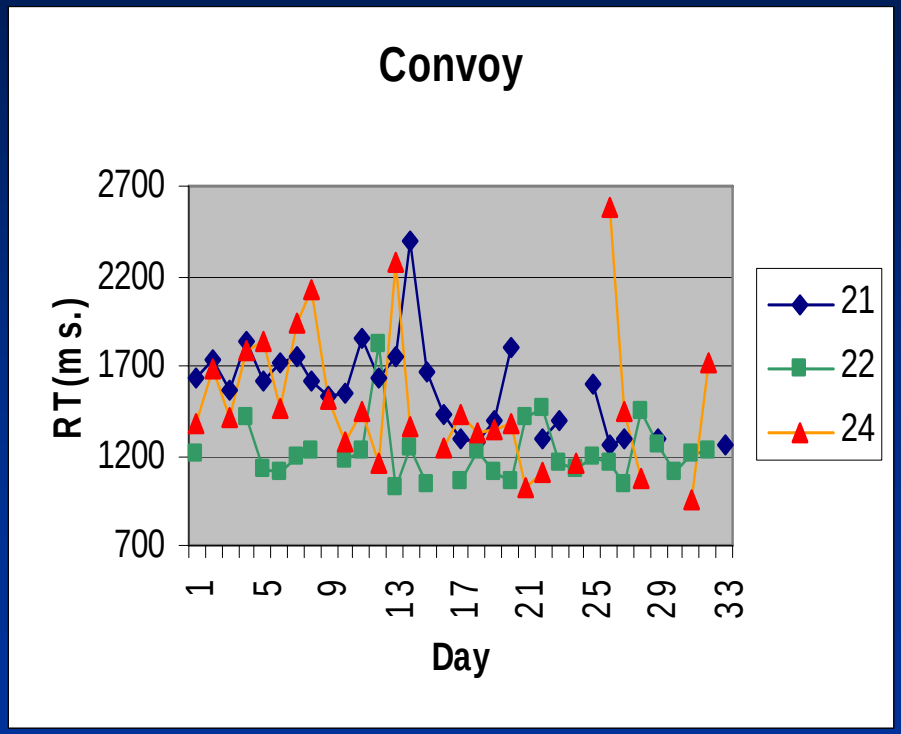
Response Time

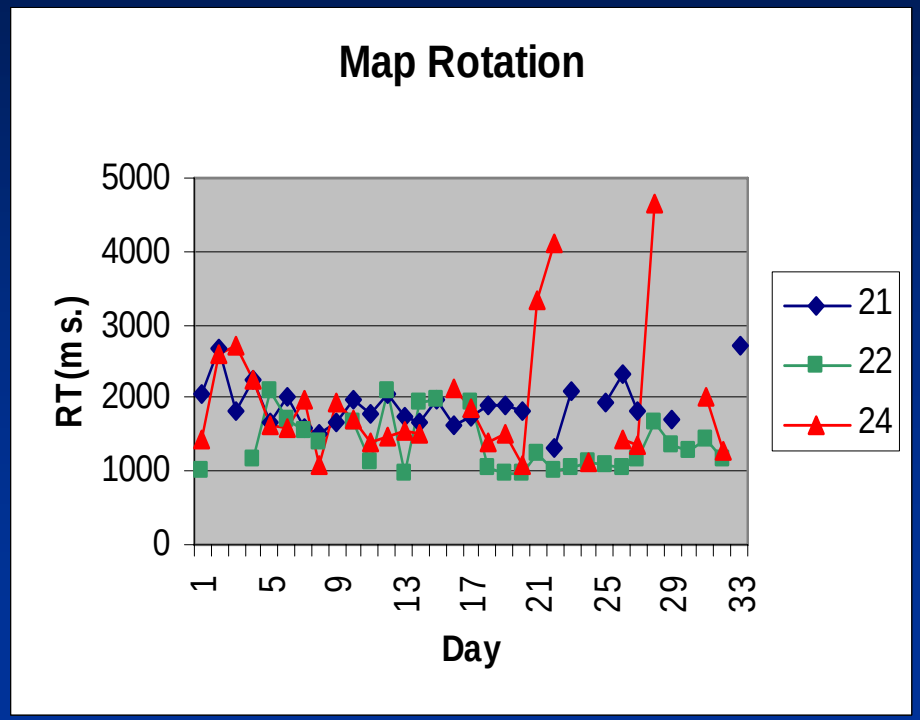
- Large individual differences
- Trends differed in individuals
- Data indicates state of person

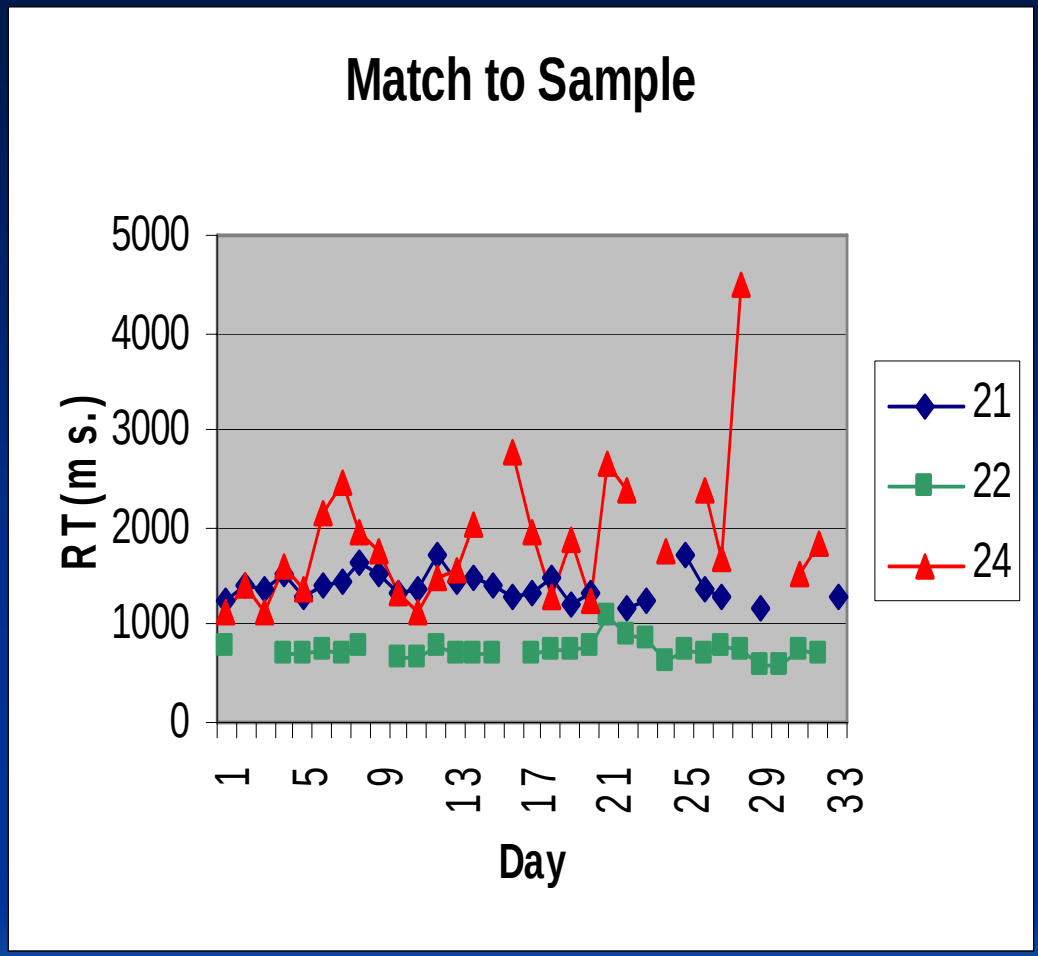



Shoot/No shoot
RT (m s.)
600
550
500
450
400
350
300
250
1
5
9
13
17
21
25
29
33
Day
21
22
24


Classify
RT (m s.)
800
700
600
500
400
300
1
5
9
13
17
21
25
29
33
Day
21
22
24

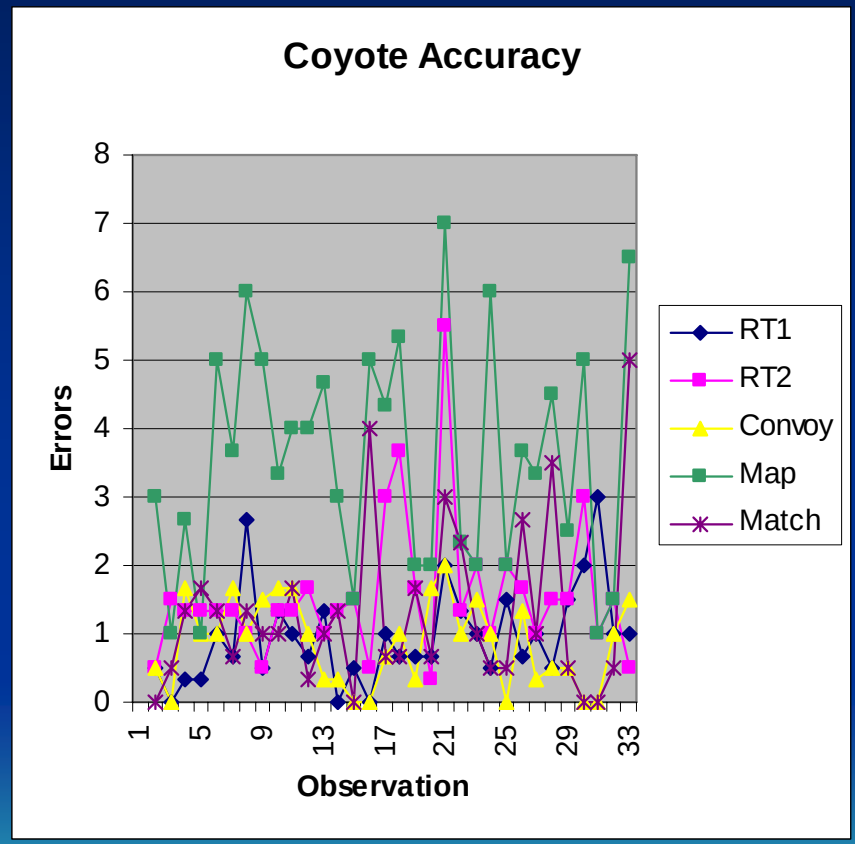

Convoy
2700
2200
1700
1200
700
RT(m s.)
1
5
9
13
17
21
25
29
33
Day
21
22
24

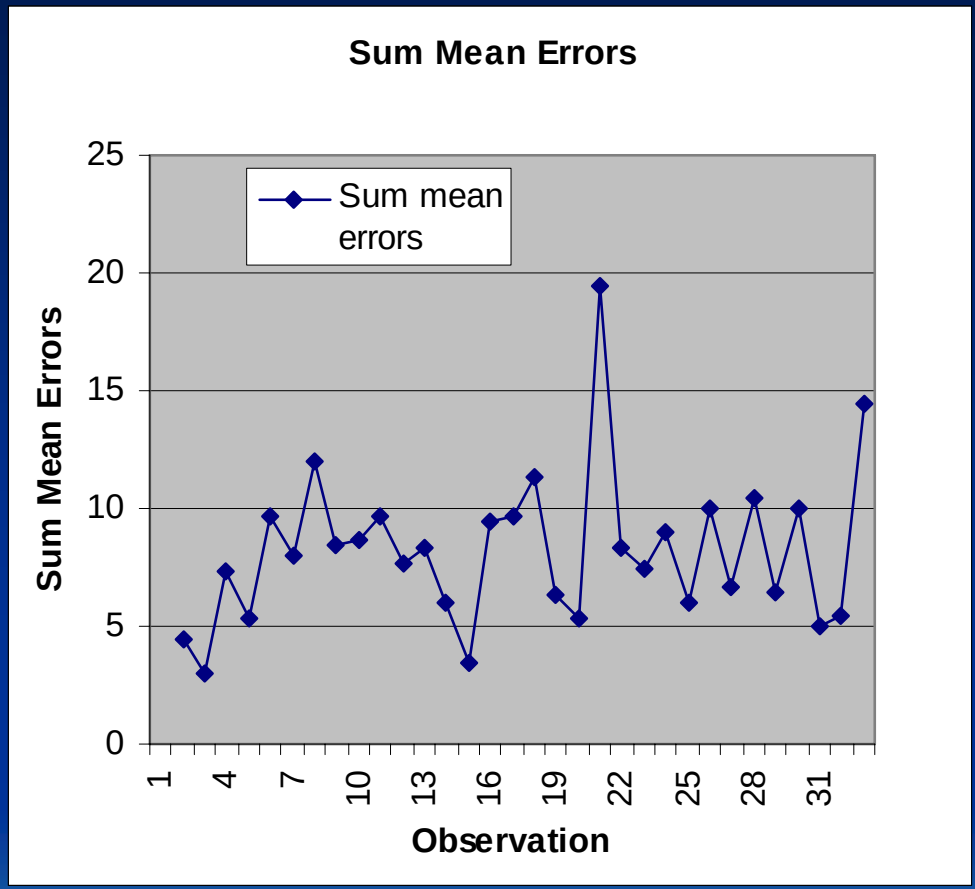

Map Rotation
5000
4000
3000
2000
1000
0
RT(m s.)
1
5
9
13
17
21
25
29
33
Day
21
22
24

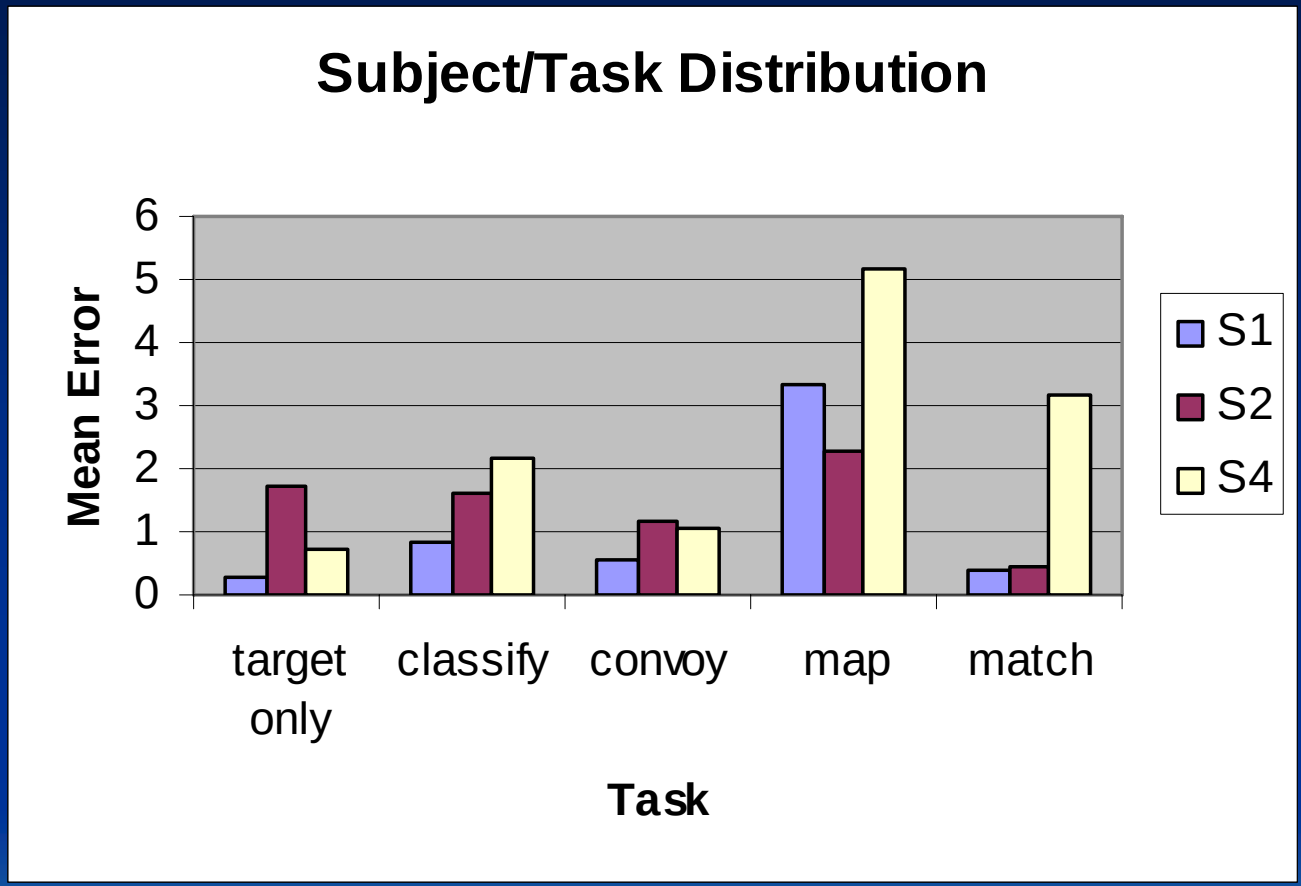

Match to Sample
5000
4000
3000
2000
1000
0
RT (m s.)
1
5
9
13
17
21
25
29
33
Day
21
22
24

Accuracy

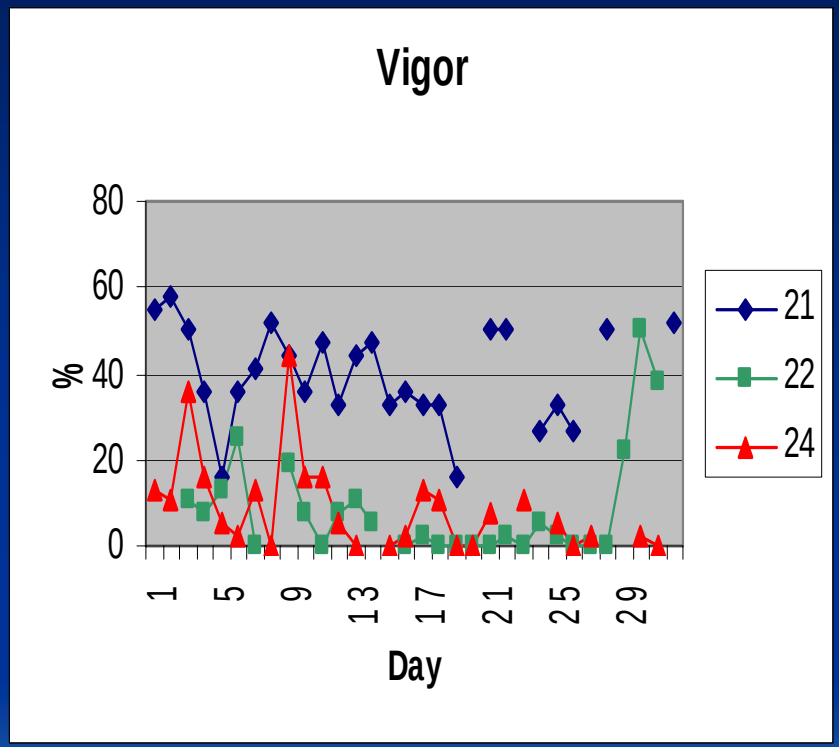
- Low error rate
- Large individual differences
- Large task differences
- Trend toward more errors
- Detection of high stress days?

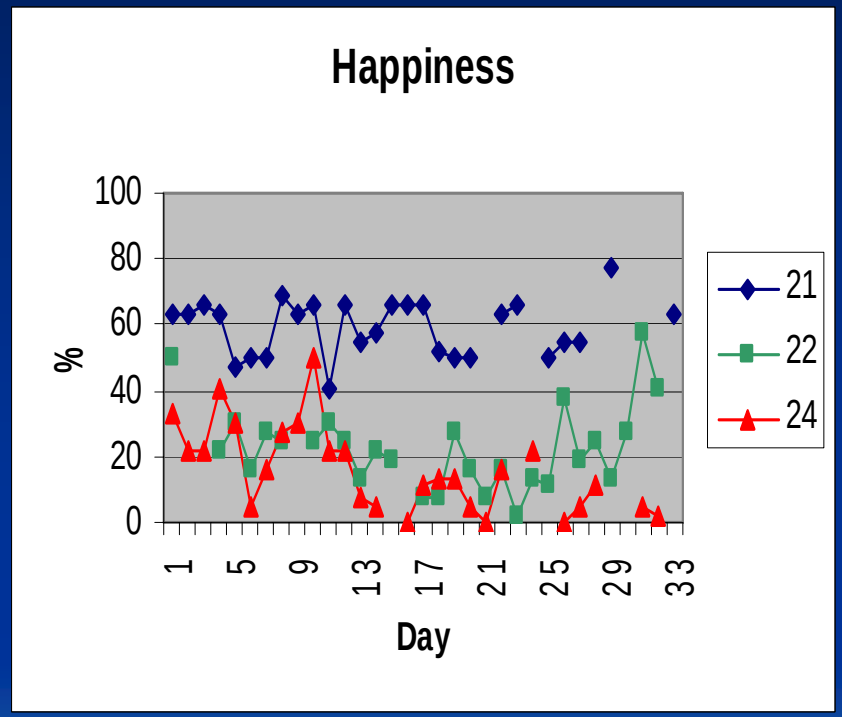



Sum Mean Errors
Sum mean errors
Sum Mean Errors
25
20
15
10
5
0
1
4
7
10
13
16
19
22
25
28
31
Observation

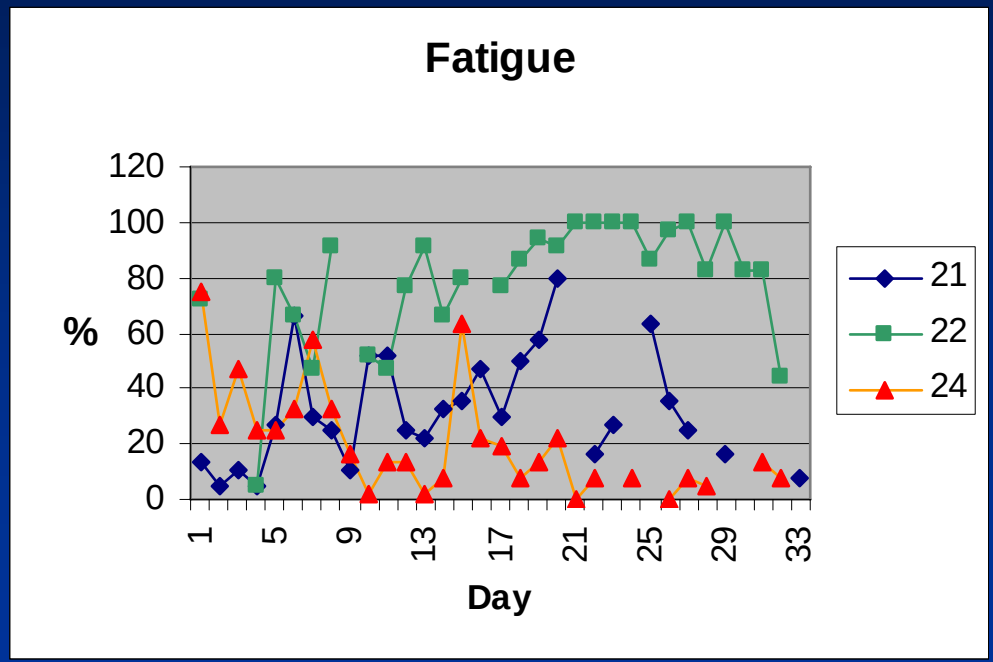

Subject/Task Distribution
Mean Error
6
5
4
3
2
1
0
target only
classify
convoy
map
match
Task
S1
S2
S4

Mood


Vigor
80
60
40
20
0
%
1
5
9
13
17
21
25
29
Day
21
22
24


Happiness
100
80
60
40
20
0
%
1
5
9
13
17
21
25
29
33
Day
21
22
24

- Large individual differences
- Trends – fatigue increased, vigor decreased
- Mood improves as end of training approaches



Participant Comments

- Time commitment a problem
- Sometimes enjoyable
- Map task – long, frustrating, tedious

Conclusions

- Large individual differences
- Stability during extended stress
- Different task sensitivity
- Large day-to-day variability

Acknowledgements

- DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office
- Space and Naval Warfare Systems Command, San Diego
- Marine Corps Air Ground Combat Center, Tactical Training and Exercise Control Group – Twentynine Palms

Vibrotactile Guidance Cues for Target Acquisition

Tal Oron-Gilad[1&2*], Joshua L. Downs[1], Richard D. Gilson[1], & P.A. Hancock[1&2]
Department of Psychology[1] and Institute for Simulation and Training[2]
University of Central Florida

Abstract—**Three experiments examined the use of vibrotactile cues to guide an operator toward a target. Vibrotactile stimulation on the hand can provide spatially stabilizing cues for feedback of subtle changes in position. When such feedback is present, a deviation from the point of origin results in tactile stimulation indicating the direction and magnitude of the positional error. Likewise, spatial deviation from a desired position displayed tactually can provide robust position guidance and stabilization sufficient to improve the acquisition time and accuracy of fine cursor control. A major advantage of this mode of information representation is that it can be present at the same time as visual cues with minimal cross-modal interference. Our findings suggest that performance is actually enhanced when both tactile and visual cues are present. Although previous studies have suggested that various forms of tactile feedback can provide position guidance and stabilization, to our knowledge, this work is the first that detailes the effect of tactile feedback on target acquisition directly.**

Index Terms—**Vibrotactile, Guidance Cues, Target Acquisition**

I. INTRODUCTION

Present combat operational forces must face frequent sniper opposition and often engage in small, troop size covert operations at night where unexpected, street to street conflicts may develop. Under such circumstances, it is crucial to know from which direction enemy fire is coming, especially to reduce friendly-fire and civilian casualties. Within this general realm of concern for target detection and identification, the purpose of the present studies was to establish how vibrotactile guidance cues can be used to improve marksmanship. The work originated in an effort to provide covert communication, navigation, and weapon aiming cues for infantrymen. The practical implications of this study can be extended beyond marksmanship alone into the whole range of target acquisition and warning domains where the visual environment is cluttered and complex.

A. Tactile Aiming Guidance System (TAGS)

Vibrotactile stimulation can provide spatially stabilizing cues for feedback of subtle changes in position [1-4]. Once such a feedback system is engaged, any deviation from the point of origin can result in tactile stimulation indicating the direction and magnitude of any undesired change. Likewise, spatial deviations from a desired position displayed tactually can provide robust position guidance and stabilization sufficient to improve the acquisition time and accuracy of fine cursor control [5-6]. TAGS may provide covert WY-Feel-IWYG aim-point adjustment that is as rapid and accurate as enhanced visual sights for aim-point adjustment. TAGS may also provide tactile feedback that can be used by the marksman to stabilize weapon aiming. Since TAGS apply their stimuli to the hands directly and are a tactile channel that may be capable of cooperating with the visual channel without interference [7], their stimuli may have greater affordances for aim-point guidance and stabilization than the iron and enhanced visual sights (e.g., Laser and Infrared optical sights). These affordances may translate into decreases in time to hit the target and decreases in number of bullets fired relative to visual aiming cues alone.

Though the literature suggests that various forms of tactile feedback can provide position guidance and stabilization, no current published research has detailed the effect of tactile feedback on marksmanship [1, 5, 6, 8]. As such, our investigation attempted to determine if tactile feedback with or without visual cues can improve marksmanship compared to visual feedback alone.

B. Research Questions

Our main purpose was thus to examine whether tactile cuing with or without visual cues will improve target acquisition. Given the potential for TAGS to provide guidance cues for target selection, several fundamental design issues needed to be addressed: Where should the tactors be placed to provide directional left and right cues? Does the perception of left and right change with tactor placement? Also, what kinds

This study was supported in part by a DARPA Grant "Covert multi-sensory feedback for the dismounted soldier", Contract Number DAAE07-03-C-L143, Col. William Johnson, PhD, Technical Monitor, R.D. Gilson, Principal Investigator, P.A. Hancock Co-Principal Investigator. The views expressed in this work are those of the authors and do not necessarily reflect official Army or Department of Defense policy.

*Corresponding Author: T. Oron-Gilad was with the MIT[2] Laboratory at the University of Central Florida., 3100 Technology Pkwy, Suite 337, Orlando 32826. Currently at Ben-Gurion University: orontal@bgu.ac.il.

of stimuli should TAGS employ to provide relative distance between aim-point and desired aim-point (on-target).

The first experiment was designed to address the issue of tactor placement on the hand, and examine whether change in directional right or left is related to tactor placement. The second and third experiments were designed to examine the vibrotactile cues themselves; i.e., generate stimuli that provide both direction and relative distance between the current aim-point and the desired aim-point. Both Experiments two and three explored the effect of relative distance cues and on- versus off-target tactile stimuli on reaction time and accuracy of target acquisition. Specifically, Experiment two examined the effect of relative distance cues and on- versus off-target tactile stimuli on reaction time and accuracy of aiming, while comparing the effect of presenting visual only, tactile only or combined visual + tactile cues. Experiment three further investigated the interaction between the gradient of distance cues and the on-target versus off-target interaction. Prior research has suggested that for a target acquisition task, the visual display is not necessary for the entire duration of the movement [9, 10]. The visual display is generally required for determining the direction of initial ballistic movement toward the target and then again for the further fine movements required for final target acquisition. Experiment three sought to examine whether similar characteristics are also found for tactile cuing. Both continuous and discrete vibrotactile distance cues were examined. The continuous condition was similar to that in Experiment two, providing a continuous, relative to distance, vibrotactile cue during the entire movement. In the discrete condition, an initial vibrotactile burst was provided to trigger the ballistic movement of the hand. Then, as the cursor approached the target, the tactile display was presented continuously as in the continuous condition. The distances chosen between discrete and continuous vibrotactile cuing reflect the end of the first ballistic movement, and the tolerance for establishing unambiguous initial movement, respectively.

II. GENERAL EXPERIMENTAL DESIGN

A. Apparatus

The experimental software ran on a 3.00 GHz Dell Dimension 8300 with the Windows XP Professional operating system. Screen and color resolution was fixed at 1024 x 768 and 32-bit, respectively. A Dell M992 18 inch monitor was used to project the visual display to the participants. A Gyration Ultra inertial mouse was plugged into the high-speed USB port on the computer and functioned like a conventional three-button mouse with a scrolling wheel. The vibrotactile tactor system included two EAI C2 tactors, a tactor driver, and a Velcro strap for positioning the tactors on the hand. The Tactor driver was connected to the PC (see Fig. 1). A static background image from UbiSoft's "Ghost Recon" depicting a virtual city scene looking across a street at a brick wall was displayed for the duration of the trials (northwest wall in map "m05_embassy.env"). (see Figure 2). A small grayed rectangle at the center of the image was where the default start point of each trial began. The visual stimuli were superimposed on this background in the visual conditions (see Fig. 3).

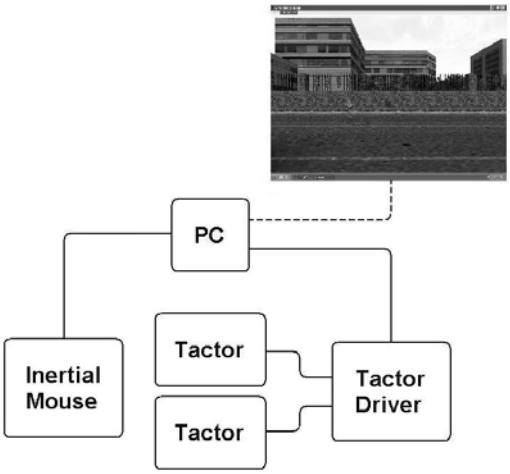


Fig. 1. Block diagram of the experimental system used in Experiments 1-3. A static background image was projected on the PC monitor; visual target cues were added when applicable.



Fig. 2. The static background image used for all three experiments. The small grayed rectangle at the center of the image is where the default start point of each trial began.

Targets: To incorporate the Fitts law paradigm [11-14] small and large targets were generated to the left and to the right of the center of the display. Targets consisted of a soldier from the Ghost Recon game holding an AK-74 pointed at the participant (actor "m05_eli_ak74_1.atr"). The target was captured in perspective for its location and size, and included a shadow. Small targets were 14 pixels wide; large targets were 28 pixels wide. The centers of mass of the target positions were located 423 pixels from the center of the display for the fart targets, and 169 pixels from the center of the display for the near targets. The order of presentation of the 8 possible targets (Right-Left, Small-Large, Near-Far) was partially counterbalanced using a sliding Latin Square technique for each participant. The mouse cursor was depicted as a white cross ('+') 19 pixels across, and always started a trial in the center of the screen. The cursor was constrained by the software to move only in the horizontal plane passing

through the center of the screen and the center of mass of all targets. Target guidance cues were provided by the visual display, by the tactile display, or by both the visual and tactile display, depending on the experimental condition. For experimental conditions where visual cues were available (in Experiments 2 and 3), targets appeared visually on the screen superimposed on the background. In the tactile only conditions, only the cursor appeared on the background image.



Fig. 3. A visual target on the background image. This particular target was a Large Near target.

Tactor placement: tactors were placed either on the palm (ventral) or on the back of the hand (dorsal). Ventral placement refers to the tactors being placed between the mouse and the participant's hand (see Fig. 4a). Dorsal placement refers to the participant's hand being placed between the mouse and the tactors (see Fig. 4b). Both positions minimally interfere with the manipulation of the mouse. Tactor placement was adjusted for each participant separately and a Velcro strap was used to secure the tactor position.

Vibrotactile stimuli: The vibrotactile stimulus was a modulated 250 Hz sinusoidal signal held at a constant gain for all participants [15,16]. Skin is most sensitive to light vibrations around 200 Hz [17], and maximum sensitivity for vibratory touch stimuli occurs from 200 to 400 Hz at stimulus intensities ranging from -20 to +60 dB [15, 18, 19]. Frequency sensitivity is rather limited and should be used with great discretion as a stimulus variable in tactile communication [20]. The IEA actuators are relatively fast, and reach the set frequency within 2-3 milliseconds which results in clear distinctive stimulus bursts. White noise was presented via headphones to mask the sound of the mechanical relays used in the tactor driver. We used burst rate as our primary source for coding as the use of temporal patterns have been successfully demonstrated before (see for example [15]). The vibrotactile burst was kept constant at 100 ms for all experiments for both directional cues and on-target cues.

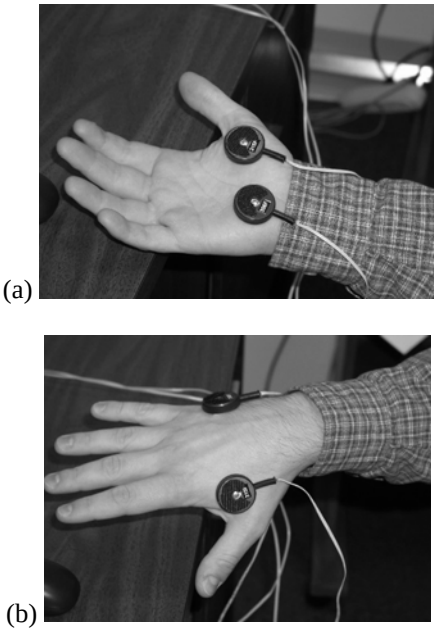


Fig. 4. Location of tactors on the hand in Experiment 1; ventral (a) or dorsal (b).

Vibrotactile directional cues: Burst rate varied among Experiments. In Experiment one the burst rate was kept constant at 6.7 bursts/sec (a stimulus burst of 100 ms and an inter stimulus interval (ISI) of 50 ms). In Experiments two and three burst rate changed gradually as a function of horizontal distance from the target. It was driven by a 3rd order polynomial function ranging from ISIs of 250 ms to 10 ms (where 10 ms is often considered as the lower limit for gap detection and see [21, 22]). Discriminability among the gradual changes of the polynomial function were not examined. Yet, discriminability between the two extremities of the function; frequent (about 9 bursts /sec, 100ms bursts and ISI of 10ms) and infrequent burst rate (about 3 bursts /sec, 100ms bursts and ISI of 250ms) was determined. Two movement profiles (gradients) of ISI changes were generated; Gradient Up in which as the cursor approaches the target the cuing becomes more frequent or Gradient Down in which as the cursor approaches the target the cuing becomes less frequent (see Fig. 5). The gradual change in burst rate was applied whenever the cursor was within 60 degrees of the target; beyond 60 degrees (which meant that the cursor was relatively far from the target), burst rate was kept constant at either frequent or infrequent burst rate depending on the gradient and the direction of the cursor (to the left or to the right of the target respectively).

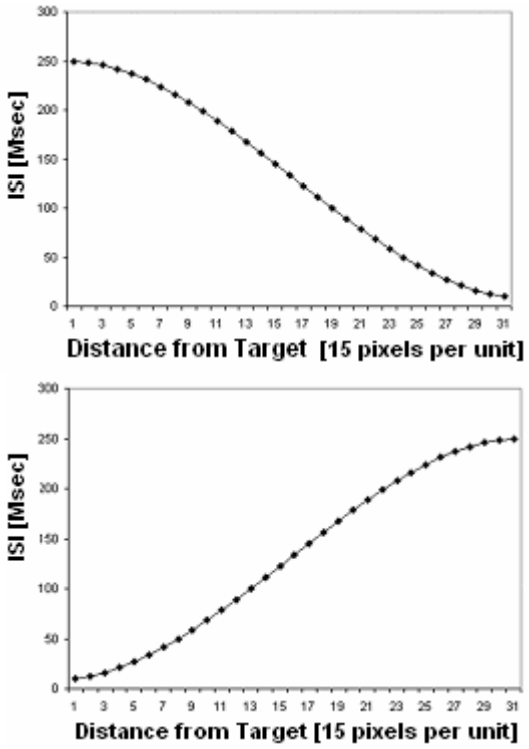


Fig. 5. Interstimulus interval by scalar distance from target used in Experiments two and three. The upper gradient profile (Gradient Down (GraD)) is one in which as the cursor approaches the target the cuing becomes less frequent. The lower gradient profile (Gradient Up (GraU)) is one in which as the cursor approaches the target cues become more frequent. Discriminability was experimentally determined between the frequent (about 9 bursts /sec, 100ms bursts and ISI of 10ms) and the infrequent burst rate (about 3 bursts /sec, 100ms bursts and ISI of 250ms).

Vibrotactile discrete directional cues: It has been suggested that visual cueing is not necessary for the entire duration of the movement of a rapid target-selection task [9, 10]. Rather, cues providing directional information must be provided during the initial movement stage and again near the target. This claim was examined for vibrotactile directional cues in Experiment three. Based on the average movement profile derived from Experiment one, two movement ranges were defined: 1) initial movement; from target pop-up to 8 pixels of movement toward the target, and 2) the end of the first ballistic movement; movements toward the target within 105 pixels of the nearest edge of the target. Those ranges were imposed on the continuous profile to create a discrete cueing condition (see Fig. 6). Movement away from the target always resulted in a continuous tactile cue with frequent or infrequent burst rate for the Up and the Down gradients respectively

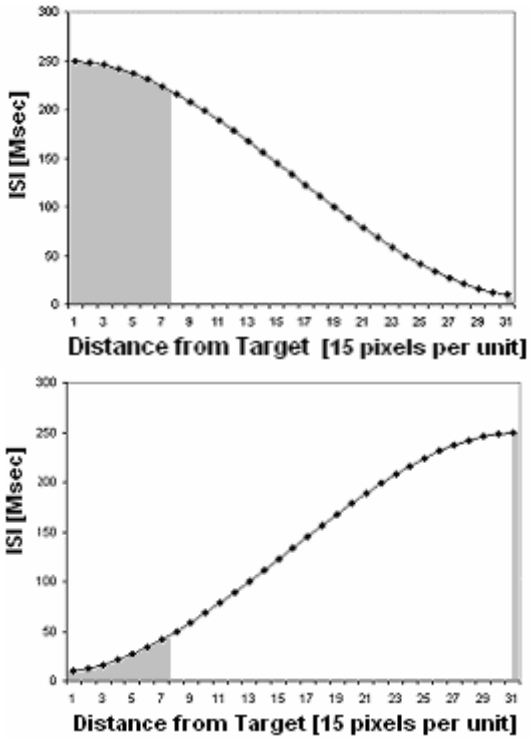


Fig. 6. The discrete and continuous profiles used in Experiment 3. The Continuous profile is similar to the one used in Experiment 2. The grayed area represents the discrete condition. Cuing occurred only during the initial 8 pixels of movement and during the last 105 pixels near the target, as the initial ballistic movement ended and fine sub movements began.

Vibrotactile on-target cues: on target cues were applied from both tactors simultaneously. Burst rate on target was kept constant but varied among experiments. In Experiment one burst rate was 6.7 bursts/sec. In Experiment two cues on target were either suppressed (silent, no cues while on target) or enhanced (frequent burst rate on target, about 9 bursts/sec). In Experiment three cues on target were either suppressed, enhanced fast with frequent burst rate on target or enhanced slow with infrequent burst rate on target (about 3 bursts/sec) depending on the experimental trial.

Tactor discrimination test: The test was administered to establish the magnitude of bias, if any, between the two tactors. Participants were presented a randomized set of left, right, or both tactor stimuli and asked to select which tactors were activated. Each stimulus burst was of 100 ms. Each combination of tactors (left, right and both) was presented 6 times, for a total of 18 trials. "left" and "right" buttons were used for indicating the left and right tactors, respectively. The "both" button was positioned in the center of the screen. Upon completion of the 18 trials, the participants shifted a continuous scrollbar left or right to indicate the relative intensity of the left and right tactors. The gain of the tactors was kept constant throughout the test.

B. General Procedure

In all three experiments the participants sat comfortably in front of the computer monitor. Tactors were placed on the hand in which they normally used a computer mouse, and the tactors were adjusted properly for each participant. Tactors were placed in line with the thumb and fourth finger at the base of the palm, either on the ventral or dorsal surface of the hand, depending on the experimental condition. The tactors were held in place by a Velcro strap wrapped around the hand. Participants had the tactors in contact with their hand for the entire duration of the experiment. Prior to any trials we made certain that participants reported one-hundred percent detection rate for the tactile stimuli for both tactors and that tactor intensity was perceived as equal from both tactors (see tactor discrimination test).

III. EXPERIMENT 1

As the locus on the skin is the primary coding mechanism for vibrotactile displays [15] it was important to establish the location of the tactors on the hand. Experiment one examined the initial response to directional (to the left or to the right) vibrotactile guidance cues for two tactor placements; on the palm (ventral) versus on the back of the hand (dorsal).

The participants hand was oriented perpendicular with respect to the floor (see Fig. 7). This position of the hand is a common grip position and as such suitable for TAGS. It is also a hand position where (for a right hand) the left side of the hand (thumb up) and the right side of the hand are not in accordance with the left side of the trunk and the right side of the trunk. It was expected however, that for a hand facing perpendicular to the ground (as shown in Fig. 7) vibrotactile direction cues applied to the left side of the hand (thumb) will result in movement of the hand to the left and cues applied to the right side of the hand (pinkie finger) will result in movement of the hand to the right regardless of the tactor placement (ventral or dorsal) since trunk position is the dominant factor in determining movement direction.

To solely examine the tactile affordances, no visual cues were presented in this experiment.

A. Experimental Design

A two-way repeated measures design with tactor placement (Ventral vs. Dorsal) and target location (Left vs. Right) as the within participant variables.

B. Experimental Participants

Twenty four undergraduate students at the University of Central Florida participated in this experiment. There were 12 males and 12 females. Though 3 males and 1 female stated that they wrote with their left hand, all participants stated that they use their right hand for mouse operations. There were no significant correlations between age, gender, mouse use in hours per day, and computer game play in hours per day with initial movement time and probability of correct initial movement.

C. Apparatus

Apparatus was as described in the General Experimental Design. The mouse was used in its inertial mouse mode such that movement of the cursor fully left or right from the center of the screen required a wrist flexion or extension of 60 degrees.



Fig. 7. Hand position and holding of the inertial mouse in Experiment 1. Note that the tactors were placed either ventral or dorsal as shown in Figure 6.

Initial movement time (iMT), probability of correct initial movement direction (iMove), and time-stamped movement profiles were collected for each trial. iMT for this experiment was defined as the time in milliseconds between target pop-up and the start of movement by the participant. iMove for this experiment was defined as the probability of making a correct initial movement toward the target by the participant. Movement profiles consisted of the time-stamped (in ms) 'x' screen coordinate of the center of the cursor recorded once every mouse tick. Mouse ticks only occur when there is movement of the mouse, with a maximum recording rate of about 100 mouse ticks per second for the described system.

D. Procedure

Participants were assigned to one of two orders of presentation of the tactor placement (Ventral or Dorsal). The participants were presented 2 blocks of 16 targets, for a total of 32 trials. Upon completion of the first 16 trials, the tactor location was switched. Before each block of trials, the tactor placement was verified by obtaining the participant's subjective perception of the discriminability and comfort of the tactors. This was accomplished by pulsing first one, then the other tactor, and having the participant point to the tactor they felt was activated.

For all trials the participants sat comfortably in front of the computer monitor with their hand unsupported (see Fig. 7). Participants were instructed to depress and hold the inertial mouse mode button under the mouse with their index finger whenever they wanted to move the cursor. Participants were not instructed to keep their eyes open during the experiment; however, we did not observe any participants close their eyes systematically throughout the experiment.

Participants practiced using the inertial mouse to move the cursor fully left, right, up, and down by using only hand motions about the wrist. Then the experiment began.

The primary task of the participants during a trial was to quickly move their hand in the direction of the target when they had an idea where the target was located. When a trial began, the tactile stimulus was presented. It continued to be present until the trial ended. Irrespective of the correctness of movement direction, each trial ended when the participant moved the cursor beyond the distance the target was located from the center of the screen. When the trial ended, the participant returned his or her hand position to neutral position and waited for the next trial to begin.

E. Results

The GLM in SPSS 11.5 was employed to analyze the two-way, repeated measures design. All tests were run at the $\alpha = .05$ level. Unambiguous intentional movements did not typically appear within 1 degree of hand movement, suggesting that the tolerance for identifying the initial movement could be widened to 8 pixels (1 degree of hand movement) rather than the 1 pixel tolerance employed by the data collection program. This 1 degree tolerance was applied to the iMT and iMove data.

The data indicated that movement to the left or movement to the right based on vibrotactile guidance cues is independent of the location of the tactors on the hand when both tactors are located on the ventral or dorsal surface. There was no significant interaction between tactor placement on the ventral (V) or dorsal (D) surface and targets left (L) or right (R) on initial movement time (M_{VL} = 648; M_{VR} = 644; M_{DL} = 709; M_{DR} = 679; F(1,23) = .966, $p > .05$, $\eta_p^2 = .040$, $\beta = .156$). There was no significant interaction between tactor placement and target left or right on iMove (M_{VL} = .776; M_{VR} = .755; M_{DL} = .734; M_{DR} = .682; F(1,23) = .195, $p > .05$, $\eta_p^2 = .008$, β = .071). When the user's hand orients perpendicular with respect to the floor (as shown in Fig. 7), the results suggest that there were no differences between placements in relation to the direction of movement. We therefore chose to use the ventral placement for the following studies as it enabled a better grip on the mouse by the participants.

IV. EXPERIMENT 2

Experiment two established the effect of continuous relative distance cues and on- versus off-target vibrotactile stimuli on reaction time and accuracy for target selection. Visual target cues were set against a visual background that had to be searched (as it is most likely to be in environments where targets are camouflaged intentionally). Tactile target cues were set against a relatively quiet background, and always correctly indicated the direction of the target. As such, it was expected that tactile target cues would facilitate target search and movement accuracy, and that using visual and vibro-tactile direction and distance cues combined would result in faster time-to-target compared to visual cues alone.

The initial tactile burst on target pop-up provided the direction (right or left). The manipulation of tactile burst rate, provided the gradient (i.e., the distance from the target). Increasing burst rate gave rapid feedback during near-target sub-movements. Decreasing burst rate gave initial cues on target popup, but less rapid feedback during near-target sub-movements. It was expected that faster tactile burst rates near the target would result in faster time-to-target compared to slower burst rates near the target as the faster burst rates close to the target provide feedback for the final approach to the target. Furthermore, it was also important to identify whether the vibrotactile cues on target should be enhanced (frequent burst rate on target) or suppressed (no pulsing on target (silent)). Subject matter experts in marksmanship have indicated that a suppressed or silent target is preferred since the extra stimulation while on target may be distracting the operator from the target acquisition task. Experiment two was designed to empirically explore this claim.

A. Experimental Design

Experiment two employed a 3 x 2 x 2 mixed factorial design. The within-participant variable was Display (visual vs. tactile vs. visual + tactile combined). Between-participants variables included cuing on target (Suppressed or Enhanced) and Gradient (Up or Down).

B. Experimental Participants

The twenty four undergraduate students at the University of Central Florida who participated in Experiment one also participated in this experiment. Hence, there were 12 males and 12 females in the sample. Each participant performed both experiments on the same day. A short break was offered between the experiments, but all participants declined. There were no significant correlations between age, gender, mouse use in hours per day, and computer game play in hours per day with initial movement time, probability of correct initial movement, and target selection time.

C. Apparatus

The setup for Experiment two was similar to Experiment one, with the following changes. The mouse was used in its optical mouse mode on the desk in front of the monitor. Hence participants no longer had to hold the mouse in their hand in the air. Practically, it was difficult for participants to hold their arm freely without support (as in Experiment one) for the duration of a longer experiment. Tactors were positioned on either side of the mouse such that the thumb and third finger were in direct contact with the tactors' vibrating elements (see Fig. 8) similar to the ventral position in Experiment 1. Movement of the cursor onto the near targets required a 2.5-inch movement of the mouse from the central point of origin. Movement of the cursor onto the far targets required a 5.0-inch movement respectively. In the visual and visual + tactile conditions visual targers appeared as described in the general experimental method (Fig. 3).

Initial movement time (iMT), probability of correct initial movement (iMove), the number of times on-target (otCnt), time from target pop-up to target selection (ST), and time-stamped movement profiles were collected for each trial. iMT for this experiment is defined as the time in seconds between target pop-up and the start of movement by the participant. iMove is the probability of a correct initial movement toward the target by the participant. otCnt is defined as the number of times the cursor went from off-target to on-target. Movement profiles consisted of the time-stamped (in seconds) horizontal screen coordinate of the center of the cursor recorded every time the mouse position changed. A maximum recording rate of about 100 mouse movements per second was achieved for the described system.



Fig. 8. Hand position using the inertial mouse for Experiments 2-3, here the hand was supported by the desk and the tactors were placed on the mouse.

D. Experimental Procedure

Participants were randomly assigned to one of four groups: target Suppressed (TarS), gradient Up (GraU); target Suppressed, gradient Down (GraD); target Enhanced (TarE), gradient Up; and target Enhanced, gradient Down. The participants were presented 3 blocks of 32 targets, for a total of 96 trials. The first block of trials was visual only, the second block of trials was tactile only, and the third block of trials was visual and tactile. Order of presentation of the blocks was not varied in this experiment. Before each block of trials, the instructions specific for the block were briefly reviewed. Upon completion of the experiment, the participants were asked to fill out a questionnaire about their experience with computers and video games, and their subjective experience during the experiment.

The primary task during a trial was to quickly move the cursor onto the target, and then, when on target, to click on the left mouse button. When a trial began, the white cross cursor appeared in the center of the screen for all conditions. In the visual conditions, the target appeared at random position in the horizontal plane. The target continued to be presented until the trial ended. Each trial ended when the participant clicked on the target. For the tactile condition, tactile cuing did not stop until the participant clicked on the target. If the participant missed the target and clicked elsewhere, tactile cuing continued to indicate that he did not hit the target yet. The next trial began after a random delay ranging from 2 to 9 seconds.

E. Results

The GLM in SPSS 11.5 was employed to analyze the 3 x 2 x 2 mixed factorial design. All tests were run at the $\alpha = .05$ level for significance testing.

There were no significant differences for selection time between the visual condition and the visual + tactile condition (M_{ST}=1423 ms (SD=241), and M_{ST}=1368 ms (SD=175). However, there was a small trend (Cohen's d=0.261) toward shorter selection times in the tactile+visual condition. The visual and visual + tactile conditions were significantly better than the tactile only condition (M_{ST}=6647 ms (SD=4080).

There was a significant interaction between the gradient of vibrotactile stimuli (GraU or GraD) and the method used to highlight the target (TarS or TarE) for selection times. The suppressed target condition (M_{TarS} = 4395 ms) was superior to the enhanced target condition (M_{TarE} = 8899 ms) in terms of target selection time (ST) ($\boldsymbol{F}(1,20) = 11.017$, $\boldsymbol{p} < .05$, $\eta_p^2 = 0.355$) (see Fig. 9). This was particularly evident when the pulse rate increases as the cursor moves closer to a target (GraU).

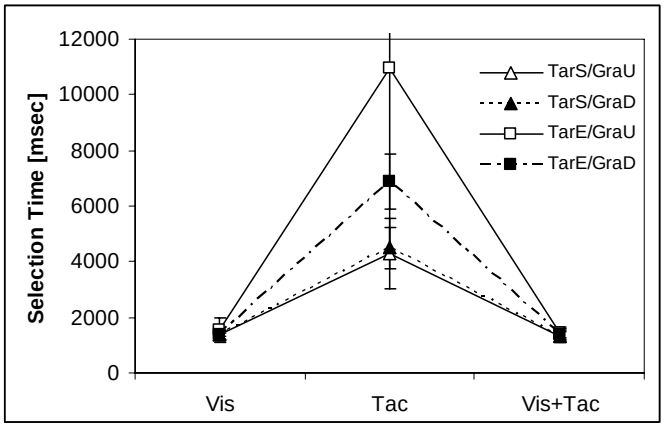


Fig. 9. Target Selection time in Experiment two where there was a significant interaction between the cuing on target (suppressed or enhances) and the gradient of the movement (up or down, and see Figure 7) .

For number of movements on target until acquisition (otCnt) there was a significant Display x Target x Gradient interaction ($\boldsymbol{F}(2,40)=3.94$, $\boldsymbol{p}<.05$, $\eta_p^2 = 0.165$). Here the tactile only condition was inferior to the visual and the visual+tactile conditions across all combinations resulting in more times to cross over the target before acquiring it as shown in Fig. 10. The tactile+ visual condition was not significantly different from the visual condition except for in the gradient down and target enhanced condition (M=1.281 (SD=0.201), and M=1.146 (SD=0.151), respectively).

With regard to initial movement time there was a significant main effect of display ($\boldsymbol{F}(2, 40) = 33.350$, $\boldsymbol{p} < .05$, $\eta_p^2 = 0.625$) indicating that the initial movements in the

visual+tactile conditions were significantly faster than in either one of the other conditions (M=404 (SD=63), M=386 (SD=83), and M=288 (SD=59) ms respectively for the visual, tactile only, and visual+tactile conditions). There were no significant differences in initial movement between the visual and the tactile only conditions as shown in Figure 11.

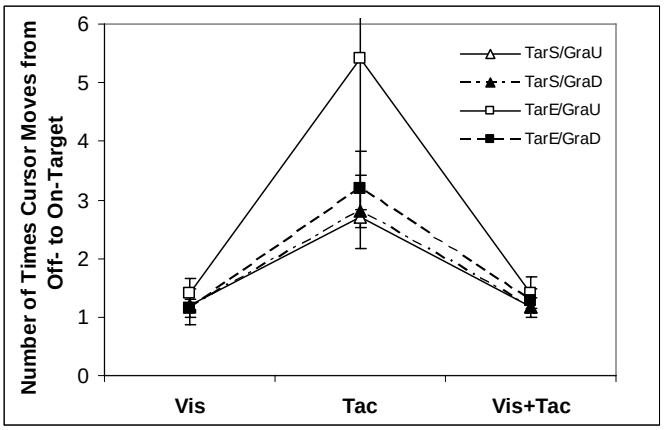


Fig. 10. Number of movements on Target in Experiment two. Here again the same interaction between the cuing on target (suppressed or enhances) and the gradient of the movement (up or down, and see Figure 7) was found .

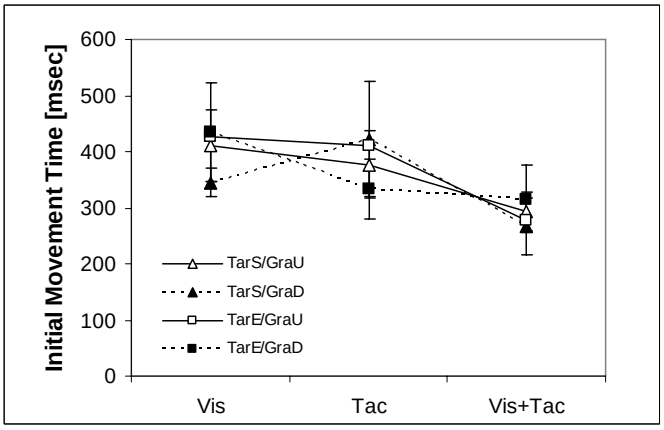


Figure 11. Initial movement times in Experiment 2 where the combined Visual+Tactile display provided the fastest responses.

Experiment two showed that there was an interaction between the gradient of vibrotactile stimuli and the method used to highlight the target. The suppressed target condition (where no tactile stimulus was present when the cursor was on target) was superior to the enhanced target condition (where perceived intensity was at its highest on the target). Experiment two also revealed a consistent superiority of the visual and visual-tactile conditions over tactile-only. This finding can be attributed to the fact that vision is the dominant source of information for the object selection task employed. There were advantages in initial movement time to the visual+tactile condition over the visual only condition; however these differences were not significantly maintained in target selection time. Possibly, the small yet significant difference in number of times the cursor went over the target (otCnt) for the enhanced target caused the initial movement time advantage to dissipate. This may imply that for the tactile stimuli to be effective for target acquisition, a large differentiation is necessary between guidance cues and on target cues.

V. EXPERIMENT 3

Experiment three was designed to establish if there are performance differences between tactile and visual + tactile displays with respect to: 1) the interaction between the near-target pulse rate and on-target cues, and 2) differences between discrete and continuous guidance cues.

Experiment two showed that the effectiveness of distance guidance cues may interact with the on-target cues when those cues use the same basic burst rate. Experiment three further investigated this interaction between the gradient of directional cues and the on-target cues. To maximize differentiation between on-target and near-target cues [23-27], three different on-target cues were used. It was hypothesized that providing higher resolution feedback near the target (ISI < 100 ms) should improve target selection accuracy compared to lower resolution feedback (ISI > 100 ms) [21]. Minimum difference between the on-target and near-target stimuli occurs when the on-target and near-target pulse rates are the same. With this in mind, using the absence of vibrotactile stimuli on-target with infrequent burst rate near the target should result in an intermediate level of difference between the on-target and near-target stimuli. It was thus expected that variation in burst rate between near-target and on-target will result in shorter time-on-target than no or smaller variations in burst rate.

Based on research in movement control [9,10] it was hypothesized that discontinuous vibrotactile direction and distance cues (discrete cues) will result in identical target selection times to continuous vibrotactile direction and distance cues.

A. Experimental Design

Experiment three employed a 2 x 3 x 2 x 2 mixed factorial design (Display by On-target Cues by Gradient by Continuous versus Discrete stimuli). Within-subjects variables include Visual + Tactile or Tactile (Display), and Target Suppressed, Target Enhanced Slow, or Target Enhanced Fast (TarSEsf).

Between-subjects variables include Gradient Continuous or Gradient Discrete (GraCDi), and ISI Gradient sweeps Up or Down (GraUDo).

Figure 12 graphically depicts the twelve possible combinations for TarSEsf x GraCDi x GraUDo. The height of the curves represents the duration of ISI, where higher points on the curves represent longer ISI. The width of the curves represents the scalar distance from the target, where the target is in the center of convergence of each set of curves. TarSEsf is represented by the blank space or mixed-color bar between the converging curves. GraCDi is represented by the

continuity of the curves; the left set of curves is Continuous, while the right set is Discrete. Finally, GraUDo is represented by the height of the curves at the center of convergence of the set of curves.

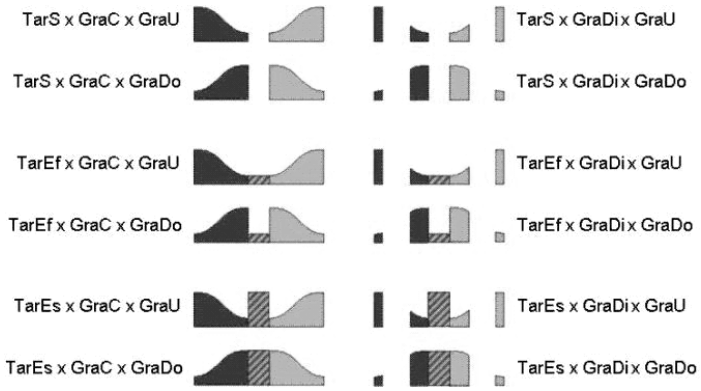


Fig. 12. Graphical depiction of the conditions in Experiment 3. The Tactile display was either continuous (GraC) (as shown in the left) or discrete (GraDi). Three levels of target cuing (Suppressed (S), Enhanced Slow (Es), and Enhanced Fast (Ef)) and two gradient profiles (Up(GraU) and Down (GraD)) generated 12 experimental combinations for each display type.

B. Participants

Thirty two undergraduate students at the University of Central Florida participated in this experiment. There were 9 male participants and 23 female participants in this sample. Participants' handedness was measured using the Edinburgh Handedness Inventory [28]. participant questionnaire collected data about their experience with computers and video games. No male participants and three female participants indicated that they have a left-hand bias, all participants chose to use the mouse with their right hand. Participants were assigned to an order of presentation of the within-subjects conditions by Latin Square. Each participant was assigned to the next order of presentation of the Display conditions in the Display Latin Square. Each participant was then assigned the next order of presentation of the TarSEsf conditions in the TarSEsf Latin Square.

C. Apparatus

The same software deployed for Experiment two was deployed for this experiment.

Initial movement time (iMT), probability of correct initial movement (iMove), the number of times on-target (otCnt), time from target pop-up to target selection (ST), and time-stamped movement profiles were collected for each trial.

D. Experimental Procedure

The procedure was similar to Experiment two, with the few exceptions. A computer-based version of the Edinburgh Handedness Inventory (EHI) [28] was administered. The tactors were then applied to the participant's preferred hand. The tactor discrimination test was then administered and the tactors repositioned as necessary to achieve zero perceived bias between the tactors.

Training before each block included a review of the stimuli that will be presented for the block, 8 training targets, and a post-training opportunity to ask questions about the stimuli.

The primary task of the participants during a trial was to quickly move the cursor onto the target and then click on the left mouse button. Each testing block included 32 trials. When a trial began, the target stimuli were presented as appropriate for the block, and the cursor appeared at the center of the screen. The stimuli continued to be presented until the trial ended. Each trial ended when the participant clicked on the target. The following trial began after a randomly generated delay ranging from 2 to 9 seconds.

E. Results

The GLM in SPSS 11.5 was employed to analyze the 2 x 3 x 2 x 2 mixed factorial design. All tests were run at the α = .05 level.

For Selection time the tactile only display resulted in significantly slower target selection times than the visual + tactile display ($\boldsymbol{F}(1,28)=274.33$, $\boldsymbol{p}<.000$, $\eta_p^2 = 0.907$) $M_{ST}=5363$ and $M_{ST}=1395$ ms respectively. No other main effects were significant. There was a significant Display x TarSEsf x GarUDo interaction (($\boldsymbol{F}(2,56)=3.375$, $\boldsymbol{p}<.005$, $\eta_p^2 = 0.118$) (see Table 1 for means). There was also a significant TarSEsf x GarUDo interaction (($\boldsymbol{F}(2,56)=3.399$, $\boldsymbol{p}<.005$, $\eta_p^2 = 0.125$) indicating that there was an interaction between the gradient of vibrotactile stimuli (GraU or GraD) and the method used to highlight the target (TarS, TarEf or TarEs) for selection time as shown in Fig. 13. The larger the difference between off- and on-target cues (e.g., TarS/GraU, TarEs/GraU, TarEf/GraDo), the less time the participant spends on the target before selecting the target, and the less time it takes from target pop-up for the participant to select the target). This appears to be particularly true when the approach to the target is with an increasing pulse rate (GraU). These results support the hypothesis that variation in pulse rate when moving On/Off the target will result in shorter time-on-target than no variation in pulse rate.

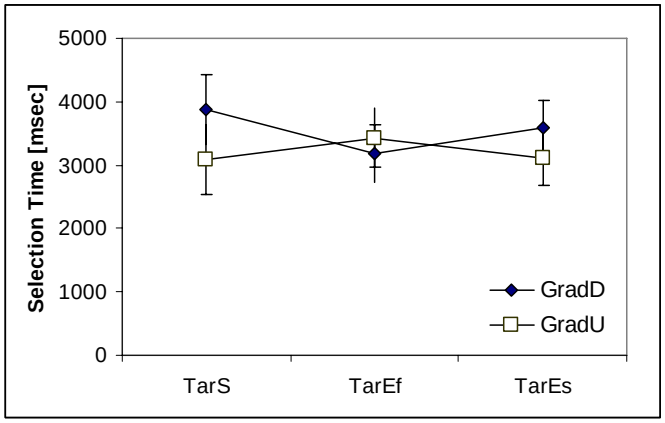


Fig. 13. Target Selection time in Experiment Three as a function of gradient and target cuing. There was a significant interaction between the cuing on target (suppressed, enhanced fast or enhanced slow) and the gradient of the movement (up or down, and see Figure 12)

Table 1. Target selection times in seconds

	Gradient	Suppressed	Target Enhanced Fast	Enhanced Slow
Visual+Tactile				
Cont.	Up	1.37	1.40	1.41
		(0.13)	(0.15)	(0.22)
	Down	1.32	1.30	1.31
		(0.22)	(0.20)	(0.19)
Discrete	Up	1.47	1.46	1.48
		(0.22)	(0.12)	(0.23)
	Down	1.41	1.43	1.39
		(0.19)	(0.21)	(0.17)
Tactile Only				
Cont.	Up	4.30	5.41	4.06
		(2.08)	(1.58)	(0.94)
	Down	5.89	5.43	5.83
		(2.39)	(2.56)	(1.41)
Discrete	Up	5.20	5.44	5.48
		(1.40)	(0.75)	(1.37)
	Down	6.91	4.57	5.83
		(3.24)	(1.44)	(1.92)

Note: Values in parenthesis are standard deviations. There were no significant differences among the discrete and continuous conditions. However there were significant differences between the Visual+Tactile and the Tactile only conditions.

For number of movements on target until acquisition (otCnt) there was a significant main effect for display ($\boldsymbol{F}(1, 28) = 125.84$, $\boldsymbol{p} < .000$, $\boldsymbol{\eta_p^2} = 0.818$) ($M_{otCn}t=1.37$ for the visual + tactile display and $M_{otCn}t=2.69$ for the tactile only display). There was also a main effect for target cuing ($\boldsymbol{F}(2, 28) = 4.351$, $\boldsymbol{p} < .018$, $\boldsymbol{\eta_p^2} = 0.134$) where the suppressed targt (no cues on target) caused more repetitions on target than either one of the enhanced conditions ($M_{otCn}t=2.16$ (SD=.33), $M_{otCn}t=1.96$ (SD=.25), and $M_{otCn}t=1.96$ (SD=.28) for the suppressed, Enhanced fast, and enhanced slow respectively). The interaction Display x Target was also significant ($\boldsymbol{F}(2, 56) = 3.29$, $\boldsymbol{p} < .05$, $\boldsymbol{\eta_p^2} = 0.105$).

With regard to initial movement time (ms) there was a significant main effect of display ($\boldsymbol{F}(1, 28) = 47.663$, $\boldsymbol{p} < .000$, $\boldsymbol{\eta_p^2} = 0.630$) indicating that the initial movement in the visual+tactile conditions were significantly faster than in the tactile only condition (M=349 (SD=88), M=484 (SD=148). These results are similar to the ones found in experiment two.

Given that there was no difference in target selection time, or numbers of time on target across the continuous and discrete conditions, the data suggest that gradient continuity is not necessary on this measure of marksmanship. The results support the hypothesis that discontinuous vibrotactile direction and distance cues will result in identical target selection times to continuous vibrotactile direction and distance cues. It should be noted, however, that in *all* cases the visual+tactile display conditions were superior to the tactile-only conditions.

VI. DISCUSSION AND CONCLUSIONS

The purpose of this study was twofold: i) to establish whether vibrotactile guidance cues can improve target acquisition over visual cues alone; and ii) to examine whether vibrotacitle cues can effectively convey not only the direction but also the magnitude of movement.

Experiment one established the effect on initial response to vibrotactile guidance cues of tactor placements on the palm (ventral) versus on the back of the hand (dorsal), and targets appearing left versus right of center. Results suggest that tactile cues provided on the left side of the medial line of the hand afford moving the hand to the left, while tactile cues provided on the right side of the medial line afford moving the hand to the right.

Experiment two established the effect of continuous relative distance cues and on- versus off-target vibrotactile stimuli on reaction time and accuracy for target selection. Results suggest that there may be an interaction between the pulse rate of vibrotactile stimuli and the method used to highlight an "on-target" condition. Generally, the suppressed target condition was superior to the enhanced target condition. This was particularly true when the pulse rate increased as the cursor moved closer to a target.

Experiment two also demonstrated the dominance of visual perception over tactile perception. When visual displays are available, the tactile cues significantly improved initial movement time. Hence, the direction of movement can be determined much faster and with more reliability when tactile cues are provided than with visual cues alone. This may be attributed to the fact that the addition of tactile stimuli reduces the need to visually search for the appearance of a new target. Particularly in the conditions simulated in our experiments there was a high degree of uncertainty regarding target onset and it most likely may be encouraging participants to adopt a strategy of attending to the tactile event cue. However, the fact that we did not find similar initial movements in the tactile only condition reinforces the notion that the tactile cues facilitate the faster movement only when the visual cues are also present. There was not a significant difference in target selection time between the visual and visual+tactile condition, possibly due to the interaction between the tactile cueing gradient and the way the target was highlighted.

Experiment three examined performance differences between discrete and continuous distance information for

target selection, and investigated the interaction between the near-target pulse rate and on-target cues. Results suggest that maximizing the difference between near-target guidance cues and on-target cues reduces the target selection time, particularly when the near-target burst rates are frequent (about 9 bursts/sec). The results also confirm that, as with vision, the vibrotactile off-target guidance cues are not necessary during the entire target selection task. Rather, the guidance cues can be provided only during the initial pop-up condition and during the sub-movements closing on the target with little or no change in performance. Futhermore, experiment three showed again, as in experiment two, that for tactile cues there was an interaction between the tactile gradient and the way the target is highlighted.

A. Practical Implications

The results obtained in this study offer some practical implications for the design of vibrotactile guidance cues for target acquisition:

1. It is possible to reduce the visual search time i.e., promote the initial movement toward a target by almost half when adding tactile cues.
2. When varying the pulse rate as a function of distance off-target, and when providing for on-target cues, the larger the difference between near- and on-target cues, the better. The worst combination of cues is to have the same on-target burst rate as the near-target burst rate. Here, we found the best combination of cues is to have a frequent burst rate (about 9 bursts/sec, e.g., ISI = 10 ms) near the target, with vibrotactile cues absent on-target. The next best combination of cues is to have a frequent burst rate near the target, with infrequent burst rate (about 3 bursts/sec, e.g., ISI = 250 ms) from both tactors on-target. (see Table 2).
3. A concern with vibrotactile sensor is that they remain fixed in place when operators move about, as in combat situation. The ventral tactor position enables positioning the tactors on the aiming tool rather then on the hand.

Table 2. Rank Order of the Best Selection Time Profiles

Profile	Selection Time (sec)	Final Time on Target before selection (sec)
TarEs x GraU	2.73	.63
TarS x GraU	2.83	.49
TarEf x GraU	2.85	.50
TarEf x GraDo	3.36	.54

B. Recommendations for Future Research

Since the tactors would most likely be applied to only one surface of the hand in TAGS (i.e., ventral or dorsal), our purpose for these studies was to establish, among other things, which surface of the hand should be employed for our given application rather than exploring the more fundamental affordance issues requiring an exhaustive analysis of the possible combinations of tactor placement and hand orientation. Such a study would permit a more complete analysis of the possible shift between negative- and positive-feedback that may occur with multi-surface tactor placement spanning a wide range of hand orientations with respect to the acceleration reference frame.

Our design focused on establishing the effect of subtle versus extreme variations in on- versus off-target guidance cues, and whether or not a continuously presented gradient is necessary for providing guidance cues to the target. This was done using a single gradient (see Fig. 7) based on analysis of movement in one study. Future research needs to explore a variety of gradients.

Experience with the tactors seems to affect participants' ability to perform the target selection task irrespective of the type of tactile guidance cues provided. From our experience we have noted that participants take the time to explore the full range of the tactile display during training rather than simply acquire the target. The effect of familiarity, experience and training in tactile displays need to be further examined.

REFERENCES

[1] A. Priplata, J. Niemi, M. Salen, J. Harry, L.A. Lipsitz, and J.J. Collins, "Noise-enhanced human balance control," *Physical Review Letters*, vol. 89(23), 2002.

[2] M. Akamatsu and I.S. MacKenzie, "Movement characteristics using a mouse with tactile and force feedback," *IJ of Hum.-Comp. Studies, vol*. 45, pp. 483-493, 1996.

[3] M. Akamatsu and S. Sato, "A multi-modal mouse with tactile and force feedback," *IJ of Hum.-Comp. Studies, vol*. 40, pp. 443-453, 1994.

[4] M. Minsky, O. Ming, O. Steele, F.P. Brooks, and M. Behensky, "Feeling and seeing: Issues in force display," *Computer Graphics*, *vol*. 24, pp. 235-243, 1990.

[5] R.J. Jagacinski, J.M. Flach, and R.D. Gilson, "A comparison of visual and kinesthetic-tactual displays for compensatory tracking," *IEEE Transactions on Systems, Man, and Cybernetics*, *vol*. SMC-13(6), pp. 1103-1983, 1983.

[6] R.J. Jagacinski, D.P. Miller, and R.D. Gilson, "A comparison of kinesthetic-tactual and visual displays via a critical tracking task," *Human Factors*, vol. 21(1), pp. 79-86, 1979.

[7] C.D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science*, vol. 3(2), pp. 159-177, 2002.

[8] J.W. Hill, "A describing function analysis of tracking performance using two tactile displays," *IEEE Transactions on Man-Machine Systems*, MMS-11, 92-1.

[9] M. Jeannerod and C. Prablanc, "Visual control of reaching movements in man," in J.E. Desmedt (Ed.) *Motor Control Mechanisms in Health and Disease*. New York, NY: Raven Press, 1983, pp. 13-29.

[10] L.G. Carlton, "Processing visual feedback information for movement control," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 7, pp. 1019-1030, 1981.

[11] P.M. Fitts, "The information capacity of the human motor system in controlling amplitude of movement," *Journal of Experimental Psychology*, vol. 47, pp. 381-391, 1954.

[12] P.M. Fitts and J.R. Peterson, "Information capacity of discrete motor responses," *Journal of Experimental Psychology*, vol. 67, pp. 103-112, 1964.

[13] P.M. Fitts and B. Radford, "Information capacity of discrete motor responses under different cognitive sets," *Journal of Experimental Psychology*, vol. 71, pp. 475-482, 1966.

[14] R.J. Jagacinski, D.W. Pepperger, M.S. Moran, S.L. Ward, and B. Glass, "Fitts' Law and the microstructure of rapid discrete movements," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 6, pp. 309-320, 1980.

[15] C.E. Sherick, "A scale for rate of tactual vibration," J. Acoust. Soc. Am. 78(1), pp. 78-83, 1985.

[16] H.A.H.C. van Veen and J.B.F. Van Erp, "Tactile Information Presentation in the Cockpit" , Proc. of the First International Workshop on Haptic Human-Computer Interaction, pp. 174-181, 2000.

[17] R. Verrillo, "Investigation of some parameters of the cutaneous threshold for vibration," *Journal of the Acoustical Society of America*, vol. 34, pp. 1768-1773, 1962.

[18] R. Verrillo, A. Fraioli, and R. Smith, "Sensation magnitude of vibrotactile stimuli," *Perception and Psychophysics*, vol. 6(6-A), pp. 366-372, 1969.

[19] M. Rothenberf, R.T. Verrillo, S.A. Zahorian, M.L. Bracman, and S.J. Bolanowski Jr., "Vibrotactile frequency for encoding a speech parameter, J. Accoust. Am., 62 (4), pp. 1003-1012, 1977.

[20] F.A. Geldard, "Some neglected possibilities of communication, Science, 131, 1583-1588, 1960.

[21] J.B.F. van Erp, "Guidelines for the Use of Vibro-Tactile Displays", Human Computer Interaction TNO Human Factors, Soesterberg, The Netherlands, 2002

[22] C. van Doren, G.A. Geschieder, and R.T. Verrillo, „ "Vibrotactile temporal gap detection as a function of age", J. Accoust. Soc. Am. Vol 87 (5), pp. 2201-2206, 1990..

[23] A. Friberg and J. Sundberg, "Time discrimination in a monotonic, isochronous sequence," *Journal of the Acoustical Society of America*, vol. 98(5, Pt 1), pp. 2524-2431, 1995.

[24] C. Drake and M. Botte, "Tempo sensitivity in auditory sequences: Evidence for a multiple-look model," *Perception & Psychophysics*, *vol*. 54(3), pp. 277-286, 1993.

[25] G. Orban, F. Van Calenbergh, B. de Bruyn, and H. Maes, "Velocity discrimination in central and peripheral visual field," *Journal of the Optical Society of America, A, Optics, Image & Science*, vol. 2(11), pp. 1836-1847, 1985.

[26] S. Stevens, "Tactile vibration: Dynamics of sensory intensity," *Journal of Experimental Psychology*, vol. 57, pp. 210-218, 1959.

[27] G. Mowbray and J. Gebhard, "Differential sensitivity of the eye to intermittent white light," *Science*, *vol*. 121, pp. 173-175, 1955.

[28] R.C. Oldfield, "The assessment of handedness: The Edinburgh Inventory," *Neuropsychologia*, vol. 9, pp. 97-113, 1971.

Work Teams in Organizations: A Historical Reflection and Lessons Learned

CHAPTER 17

Eduardo Salas, Heather A. Priest, Kevin C. Stagl, Dana E. Sims, & C. Shawn Burke

Department of Psychology

&

Institute for Simulation & Training

University of Central Florida

3100 Technology Parkway

Orlando, FL 32826

June 30, 2005

DO NOT CITE OR CIRCULATE WITHOUT PERMISSION OF THE AUTHORS

Salas, E., Priest, H. A., Stagl, K. C., Sims, D. E., & Burke, C. S. (in press). Work teams in organizations: A historical reflection and lessons learned. To appear in L. Koppes (Ed.), *Historical perspectives in industrial/organizational psychology* (chapter 17). Mahwah, NJ: Lawrence Erlbaum Associates.

Work Teams in Organizations: A Historical Reflection and Lessons Learned

Historically speaking, the focus on organizational teams has ebbed and flowed in synchronicity with ongoing changes in our world. When needed by organizations or society in general, research on teams and the phenomena that surround them are the focus of frantic observation and exploration. However, when attention is once again redirected toward societies' latest pressing problems, concern for teams is minimal at best. This pattern has played itself out time and again over the last century as team research has been driven by highly publicized tragic events (e.g., WWII, Florida airlines crash into the Potomac River, the 9/11 tragedy, Columbia Shuttle disaster), organizational changes (e.g., flattened hierarchies, distributed human resources, self-managing work teams), and technological innovations (e.g., computers, collaboration software, networks) but is, by comparison, relatively dormant the rest of the time.

This chapter illustrates the above noted repeated waves of team research, separated by relative periods of calm. In order to illustrate these waves we have structured this chapter around the events, topics, and people that have markedly contributed to the science of teams. While the information provided in this historical reflection is organized by decade and centered around key topics, we also employ a zeitgeist approach (i.e., the naturalistic theory of history). The zeitgeist approach contends that discoveries and new ideas do not develop in a vacuum and therefore must be interpreted in light of the prevailing "spirit of the times" (Boring, 1950). With this approach in mind, we have organized this review of team research in the 20th century in order to more fully capture the essence of the key actors and events as they relate to teamwork.

We begin our discussion with the early 1900's. However, this was not the first time in history a 'revolution' had generated interest in collectives. Previous marked advances in stone tool making had already led to a dependence upon hunting parties, while agricultural innovations led

to the wide spread use of farming groups. In fact, in retrospect the opening act of the twentieth century did not even turn out to be the most important period during which interest in collectives was spurred on by fundamental societal changes. Certainly, a strong argument could be advanced that the current technological revolution holds more promise for understanding the nature of teams, and for reaping the benefits of teamwork, than during any previous historical period.

We begin this chapter by reviewing team research that occurred during the first thirty years of the twentieth century. We then proceed to structure our review around successive twenty year periods. Within these time frames, the events and people that exemplify team research and practice are presented and lessons learned follow each section. Reflected in our choice of content for inclusion is the belief that while the first 60 years of the twentieth century was an important time period for teams, an even stronger wave of studies addressing team phenomena has occurred during the last four decades, and especially during the last 25 years (see for example, Beyerlein, Johnson, & Beyerlein, 2003; Campbell & Kuncel, 2001; Campion, Medsker, & Higgs, 1993; Cohen & Bailey, 1997; Edmondson, 1999; Fleishman & Zaccaro, 1992; Guzzo & Dickson, 1996; Hackman, 1990; Ilgen, 1999; Kozlowski & Bell, 2002; Marks, Mathieu, & Zaccaro, 2001; McGrath, Arrow, & Berdahl, 2000; Salas, Stagl & Burke, 2004; Sundstrom, McIntyre, Halfhill, & Richards, 2000; Swezey & Salas, 1992; West, Tjosvold, & Smith, 2003; Zaccaro, Rittman, & Marks, 2001). Consequently, the majority of this chapter is dedicated to recent programmatic research initiatives investigating team phenomena undertaken within the last 40 years. Figure 1 provides a timeline of notable events to help frame the current chapter.

Groups or Teams?

An important issue at the outset of this chapter concerns the differences between the terminologies used in the group and team literature at different times throughout history. While

examining the historical evolution of work teams in organizations, the authors were faced with a dilemma about the appropriate breadth for this review. In other words, should we limit the scope of this chapter to cover only what is termed work teams or do we adopt a more inclusive stance by reviewing all group and team research over the last one hundred years? This dilemma is indicative of the larger ongoing debate about the similarities and differences between groups and teams. This issue has been addressed in numerous other publications (see Fleischman & Zaccaro, 1992; Guzzo & Dickson, 1996; Guzzo & Shea, 1992; Katzenbach & Smith, 1993; Salas, Dickinson, Converse, & Tannenbaum, 1992) and forging an ultimate resolution is beyond the scope of this chapter. Instead, we define both terms to clarify their use within the present chapter.

Work groups are defined herein as a (1) social entity, (2) nested within a larger social and technical system, (3) comprised of members that may work interdependently to complete shared tasks, (4) whose outcomes affect others in the larger system, as well as each other (Guzzo & Dickson, 1996). *Teams* are conceptualized to be similar but different from work groups. Specifically, work teams are defined herein as complex entities characterized by: (1) two or more individuals, (2) who interact socially, (3) dynamically, (4) recursively, (5) adaptively, (6) have shared or common valued goals, (7) hold meaningful and high levels of task, feedback, and goal interdependencies, (8) are often hierarchically structured, (9) have a limited life-span, (10) whose expertise and roles are distributed, and (11) are embedded within an organizational / environmental context that influences and is influenced by enacted competencies and processes, emergent cognitive and affective states, performance outcomes, and stakeholder judgments of team member and team effectiveness (Salas, Stagl, Burke, & Goodwin, in press).

We believe teams develop from groups based on shared commitment and “strive for synergy among members” (Guzzo & Dickson, 1996 p. 309). One key differentiation between groups and

teams is that the former have individual accountability, whereas, the latter have both individual and mutual member accountability (Katzenbach & Smith, 1993). This conceptualization of teams reflects the sentiment that simply putting individuals in a group does not mean they constitute a team (Katzenbach & Smith, 1993; Salas, Bowers, & Cannon-Bowers, 1995).

In this chapter, the terms *teams* and *groups* are used interchangeably for a number of reasons. First, much of the earliest research investigating collectives was conducted by social psychologists and this literature addresses "groups" (Levine & Moreland, 1991). In contrast to initial efforts to understand such collectives, more recent research has been undertaken from an organizational psychology perspective and predominantly emphasizes "teams" (Ilgen, 1999). Therefore, the term "groups" is more appropriate for most of the research and practice described at the beginning of this chapter, while the term "teams" is more appropriate for research described in the later time periods reviewed. In fact, team research grew out of group research, so there is a natural, temporal progression. We acknowledge that there may be degrees of difference between work teams and groups. However, we will not dwell in these distinctions rather, throughout this review, as noted, we use the terms interchangeably. Our goal in this chapter is to provide a rich picture of the historical development of the theory, research, and practice centered on collectives.

Early theories and research on groups: 1900-1920's

Enduring Themes

In the initial decades of the 1900's, researchers predominantly trained in psychology (e.g., social, clinical, experimental), engineering, and management, began publishing their theories on group phenomena. This movement coincided with the burgeoning industrial revolution ongoing in the U.S. and worldwide.

The three decades from 1900 to 1930 is, however, a valuable resource for team researchers and practitioners. This thirty year period was the first time in history that group issues were addressed on a large scale using the scientific method. In their quest for efficiency (i.e., Efficiency = Outputs / Inputs), the supporters of scientific management had successfully lobbied key stakeholders of the necessity to govern their private and public sector institutions via the rational analysis of data in lieu of relying upon intuition or rules of thumb (Shafritz & Ott, 2001). The resulting body of knowledge provided a necessary but insufficient set of circumstances to facilitate organizational administration. The missing ingredient was a set of frameworks from which to interpret the collected facts and this key element was advanced by the great classical organizational theorists (e.g., Webber, Taylor, Barnard) at the dawn of the twentieth century.

Admittedly, most of the classical theorists did not start out with the intention of directly addressing group or team issues, per se. However, their theories ultimately had profound implications for individuals, groups, organizations, industries, and entire economies. In fact, the two most enduring themes of the classical period (i.e., coordination, cooperation) are both directly relevant to the core of teamwork. Coordination concerns the ongoing balancing act between horizontal differentiation (i.e. division of labor) and vertical integration (i.e., hierarchy). Cooperation is grounded in the mutual interests of subordinates and their supervisors for the greater good of the enterprise.

The classical thinkers were the first to espouse the essential role of the behavioral sciences in achieving coordination and cooperation by proactively managing organizational structure and subordinate attitudes, motivation, and actions (Shafritz & Ott, 2001). Despite the backlash against classical organizational theory by subsequent schools of thought, both coordination and cooperation remain enduring themes into the twenty-first century. In fact, coordination and

cooperation are so prominent that they are repeatedly reflected in the team theory, research, and practice presented in this section, as well as, throughout the remainder of this chapter. The contents of this retrospective examination supports our belief that modern theorists, researchers, and practitioners would be well advised to ground their initiatives in the lessons learned over the hundred years of group and team history.

Early Group Theories

One example of early twentieth century research is found in the work of French sociologist and philosopher Emile Durkheim (1895). Durkheim's research took a gestalt approach in examining how divisions of labor affect how people work together. Specifically, Durkheim (1895) observed that a group is more than a sum of its parts (i.e., more than the sum of its individual members). This insightful finding and its wide ranging implications helped propel a century of team research and led Durkheim (1895) to state "It is a condition of the group repeated in individuals because it imposes itself upon them. It is in each part because it is in the whole, but far from being in the whole because it is in the parts" (p. 56).

Unfortunately, the years before the First World War yielded little in the way of unequivocal results regarding groups. However, with the onset of WWI the world and its inhabitants changed dramatically. The United States, as well as its allies, engaged in an unprecedented united effort both on the war front and at the home front in order to meet the previously unimaginable challenges that surfaced almost daily. Although teams were undoubtedly central to this effort, little, if any, time could be spent on the systematic exploration and documentation of team issues.

In addition to the fundamental changes going on in the world, psychology itself was experiencing a paradigm shift. After John Watson's 1913 behaviorism manifesto *Psychology as the Behaviorist Views It* (Watson, 1913), psychologists' focused on observable individual

phenomena. This widespread preoccupation with the observable aspects of individuals led to less interest in and acceptance of group studies often grounded in social psychology. In fact, Floyd Henry Allport, a prominent social psychologist, remarked about his reception at Harvard University after WWI "I found Behaviorism ascendant and rampant. I found that though my Social Psychology had enjoyed before the war a much larger vogue than I had realized, it and I were now back-numbers, relics of a bygone and superseded age... Another difficulty which I had not foreseen was that the numerous graduate students...with very few exceptions...had been taught some form of mechanistic psychology, with the consequence that they looked upon me and my outlandish theories with suspicion..." (as cited by McDougall, 1930, p. 213).

Before his death in 1916, Hugo Munsterberg had successfully lobbied Allport to refocus his dissertation research on groups. Allport accepted the challenge and emerged as an influential group researcher who reported his theories as part of his widely acclaimed 1924 book *Social Psychology*. In contrast to Durkheim's research, Allport (1924) proposed that a group was only an "illusion" and that the actions or performance of a group was the same as the actions or performance of the individuals comprising the collective. While this theory starkly opposed Durkheim's position advanced decades earlier, and much of what is accepted about teams today, it was nonetheless important research in its own right. It, after all, was one of the few initiatives that systematically investigated group issues in an environment not yet friendly to research.

Other research undertaken in the first three decades of the twentieth century also produced evidence that group performance could *hinder* productivity. For example, the results from experiments such as the "Ringelman rope-pull" suggested that group force on a rope-pulling task was less than the sum of the force expected when group members pulled alone (Moede, 1927). Over the next 40 years, researchers replicated these results (Ingham, Levinger, Graves, &

Peckham, 1974; Steiner, 1972). This later research revealed that ‘process losses’, or what we call today “performance decrements” were largely due to a lack of member motivation and coordination. This interpretation is consistent with the fact that the groups assessed in Moede’s studies were comprised of members with little or no history or investment in the group. Moede’s programmatic research, and others following it in the same vein, eventually led Steiner (1972) to conclude that actual group productivity equals potential group productivity minus processes losses plus process gains (i.e., AP = PP – PL + PG).

Key Contributors

F.W. Taylor. As noted above, the first three decades of the twentieth century was a relatively quiet time in team research. However, during these thirty years there was a building wave of support for the tenets of Taylorism (i.e., scientific management), which focused on achieving valued organizational objectives by maximizing the efficiency of the individual worker (Taylor, 1911). Taylor’s approach to organizing industry did not emerge in isolation; in fact, it was heavily influenced by his friend and predecessor Henry Towne and by Adam Smith’s seminal work on modern economic thought, *The Wealth of Nations*.

What is ironic is that scientific management, with its predominant focus on individuals, ultimately spawned the development of theories that influenced how groups work together. Upon closer inspection, however, perhaps this turn of events is not so surprising given that the fourth principle of scientific management centers on fostering teamwork. Specifically, Taylor suggests that the success of an organization hinges on a re-division of work responsibilities in order to create tighter interdependencies and thereby teamwork between management and workers. Taylor describes this new arrangement as a ‘democracy’ or ‘co-operation’ (Taylor, 1911).

Mary Follett. The principles of participation rest on the premise that leadership is granted by those who are governed. This same line of reasoning has been repeatedly articulated by the likes of John Locke, Jean Rousseau, and the authors of the Declaration of Independence (McCarthy, 2002). By the 1920's, however, the practice of participative management began for the first time in history to receive systematic scientific inquiry. Organizational scholars such as Mary Parker Follet (1926) wrote a plea for participation. Although Follett's research appeared in the same historical period as Taylorism, her ideas, and the assumptions underlying them, starkly differed from the classical thinkers. In fact, coupled with the research undertaken by the Mayo team in the Western Electric Co. Hawthorne studies (see Roethlisberger, 1941), Follett's ideas represent some of the original statements comprising the organizational behavior / human resource paradigm of organizational theory.

Follett's plea for participation focused on three core thrusts: (1) collective responsibility, (2) cross-functioning, and (3) a system of committees whose purpose was integration (Beyerlein & Porter, 2000). Follett proclaimed that positive attitudes could be fostered if both employees and employers cooperated to discover the law of the situation. Once identified, the law could be dictated to either party by the other. This line of thinking emphasized a participatory approach to leadership and resonated with stakeholders at all echelons in the system. Furthermore, other ongoing research (see Lewin, 1939) paralleled Follet's ideas. This second stream of research suggested leadership is a function of the situation and the organizational environment within which an individual is embedded, thus emphasizing the importance of context (Bavelas, 1960) and continuous feedback for enabling free will and self-direction (Argyris, 1970; Drucker, 1992).

However innovative Follett's ideas were, her views were largely ignored at the time. Perhaps this is not surprising given the subjugated status afforded to women in the early twentieth

century. Furthermore, the milieu of the times was characterized by an escalating effort to fuel an ever expanding hedonistic society (i.e., “roaring twenties”). Fortunately, a new breed of scholars (e.g., Follett, Roethlisberger, Janis), most closely aligned with the emerging organizational behavior / human resource paradigm of organizational theory, advanced innovative ideas and via their persistence reshaped the organizational landscape. The work of Follett as detailed above, as well as research described in forthcoming sections of this chapter, collectively served to shift the accepted paradigm from a state of worker dependence upon management to a state of codependence (Shafritz & Ott, 2001).

Some early lessons learned:

Lesson Learned 1: Emphasis on collective behavior has been driven by repeated waves of historical revolution including those made in farming, manufacturing, and technology.

Lesson Learned 2: Group issues can be studied scientifically (Roethlisberger, 1941).

Lesson Learned 3: Groups can be beneficial (Durkheim, 1895), but sometimes they are not and result in process losses (Allport, 1924; Moede, 1927; Steiner, 1972).

Lesson Learned 4: The “whole is more than the sum of its parts” in groups (Durkheim, 1895).

Lesson Learned 5: Context matters. That is, members of organizations, and by extension groups, are influenced by their operational environment (Lewin, 1939; Bavelas, 1960).

Lesson Learned 6: Optimizing human resources is equally important to achieving organizational objectives as making the most of material resources (Taylor, 1911).

Important developments in the 1930’s and 1940’s

The Scanlon Plan

During the economic chaos of the great depression organizations became increasingly cognizant of the business necessity to maximize all of their available material and human

resources. For example, CANCADE Ltd., a manufacturer of agricultural and construction equipment, describe on their website how they survived the Great Depression by staying lean and avoiding the expenses incurred by typical manufacturing bureaucracies at the time. Regardless of their industry, mission, or strategy, organizations were highly concerned with trying to make a dollar out of fifteen cents.

In some cases, maximizing resources meant a single-minded focus on slashing costs that seems to parallel similar industry approaches over seventy years later during the economic recession that marked the early twenty-first century. Some circles, however, slowly became aware of the need to simultaneously cut costs and increase productivity in order to grow revenue and meet other established objectives. Intertwined with this new emphasis on increasing productivity was a growing appreciation for the value of human resources to achieving growth and revenue targets. Furthermore, there was a greater appreciation for treating individuals with respect and dignity not just as human capital.

It was during these tough times that Joe Scanlon devised a plan to steer through the fiscal crisis. Scanlon, a union leader in the steel industry, forged cost saving / sharing plans for both workers and management. The success of the Scanlon Plan led to the process being benchmarked and implemented across many other organizations and is largely credited with keeping many companies afloat during these turbulent economic circumstances (Frost, 1974). In fact, derivatives of this approach were found to be effective even after the Great Depression ended. Perhaps the most unique aspect of Scanlon's approach was its firm commitment to participative decision-making. It demonstrated that groups could be a source of intrinsic motivation to group members while simultaneously enhancing organizational profitability through efficiencies and productivity gains (Frost, 1974).

Team Based Selection Boards

'The war to end all wars' (i.e., WWII) changed the face of the political, economic, and social landscape of the twentieth century and beyond. One result of this turning point is witnessed in the increasing number of research studies and industry interventions focusing on and including teams. An example of this shift is found in Britain's War Office Selection Boards (WOSBs). A given WOSB was comprised of a president, a psychiatrist, a psychologist, and military testing officers. The psychiatrists and psychologists who set up and comprised the WOSBs were recruited heavily from the Tavistock Institute of Medical Psychology, which came to prominence during the 1940's.

Although initially each member of the WOSB council was responsible for a specific piece of the holistic assessment process, over time this fragmented approach "...violated the sensibilities of the assessors who valued the rounded judgment based on observation of the total person" (Highhouse, 2002a, p. 368). Eventually, WOSB assessors were combined into an assessment team that worked together as an interdependent assessment unit to gauge candidate potential. Although the interdependencies of these teams were not as high as typically witnessed in military teams, such as action teams and command and control teams, assessors did regularly meet to exchange information and reach overall judgments about a candidate's effectiveness.

Benchmarking Britain's success with the WOSBs, the Office of Strategic Services (OSS) implemented assessment teams in the United States. The OSS utilized teams of assessors to interrogate, interview, and assess would-be spies and saboteurs. One goal of this process was to gauge how a candidate would respond to the intensive pressures and possible torture that would occur if one were captured behind enemy lines. Derivatives of this process still exist today and

have been instituted by many other countries with advanced military and intelligence infrastructures.

The use of assessor teams by the WOSB and OSS to meet the challenges posed by WWII is far from the only or even most important use of teams in the military. Indeed, teams have stood the test of time and their use in the military to perform complex missions continues to increase today, perhaps largely so because teams are an intuitive work unit (Salas & Cannon-Bowers, 2000). In the decades following the World Wars, team research and practice began to accelerate, albeit slowly at first, and this movement will be discussed in greater detail in the next section.

Movement to Create Applied Research Institutes

Research Center for Group Dynamics. Kurt Lewin's research has profoundly affected the domains of leadership, experiential learning, action research, and group dynamics. In 1944, Lewin formed the Research Center for Group Dynamics at the Massachusetts Institute of Technology (MIT) (Beyerlein & Porter, 2000). The MIT center was chiefly concerned with group productivity, communication, social perception, intergroup relations, group membership, and training group leaders. After four years at MIT, Lewin's research center relocated to the University of Michigan as a division of the Institute for Social Research. In conjunction with the Office of Naval Research, Lewin and colleagues also laid the groundwork for the National Training Laboratories, which will be addressed in greater detail later in this chapter.

Lewin continues to be a powerful force that has impacted many areas of scientific study (e.g., management theory, organizational behavior, organizational development) via a cadre of former students that carry on his message (Tedford & Baker, 2000) and applied research institutes that "bridge the gap between basic research and the practical problems for which society needed answers" (McKeachie & Brim, 1984, p. 1254). Furthermore, Lewin's theory, research, and

findings helped encourage industry to adopt participative management and other similar humanistic practices (McCarthy, 2002).

Lewin's early concept of "social climate", derived from his work on emergent leadership in small groups (clubs) of adolescent boys (see Lewin, Lippitt, & White, 1939), is credited by Fleishman (personal communication, 2005), to have influenced the design of his earlier research on the effects of "leadership climate" at the International Harvester Company (Fleishman, 1953). This is notable because Fleishman's research was the first show how leadership climate, emerging from the behavior and attitudes of plant managers, influenced the nature of lower level manager's interactions with their work groups and even overrode the effects of lower manager's leadership training.

Tavistock Institute. The second World War also led to another global phenomenon, the large scale exchange of resources across borders. This change is reflected in the programmatic research of Trist and Emory. Specifically, WWII brought together a number of talented psychologists, psychiatrists, and anthropologists through the British Army (Trist, Emory, Murray, & Trist, 1997). This group of scholars was known as the Tavistock Institute and is credited with generating many innovative social ideas. In 1946, through a grant from the Rockefeller Foundation, the Tavistock Institute of Human Relations was founded in London. Their goals included understanding the health and well-being, conflict, and breakdown of individuals and organizations. Researchers at the Tavistock Institute viewed organizations as open, sociotechnical systems, in that "any system (biological or business organization) must continually interact across boundaries with its environment to obtain inputs and to discharge output" (Beyerlein & Porter, p. 9, 2000).

Tavistock researchers conducted studies on a myriad of topics including the "longwall method" which introduced some automation to the process of cutting and blasting coal in the 1950's in the form of conveyor belts. While expected to improve efficiency, the method significantly altered the nature of the roles, interpersonal relationships, and social structure of the coal mining teams and often led to a complete break down in production.

A hard lesson from the Tavistock coal mining studies speaks to the implementation of technology from a purely engineering perspective. When organizations are primarily preoccupied with parts instead of people, the implementation of technology can backfire and ultimately damage the social relationships of team members to a degree that a machine can no longer be used efficiently. As organizations have grown increasingly dependent upon technology, the lessons learned from the Tavistock Institute have generalized far beyond coal miners. In fact, the importance of fostering positive social relationships within teams continues to be an issue of concern well into the 21st century.

The Tavistock Institute also brought about the examination of groups by clinicians. For example, Bion's (1948) psychoanalytic theory of groups which addressed group attitudes and beliefs. This theory proposed that any one work group consists of two belief systems: (1) the work group belief system (i.e., performance oriented and cognizant of the group's purpose) and (2) the basic assumption belief system (i.e., dysfunctional latent assumptions that interfere with group performance). The basic assumption belief system is further characterized by dependency and unconscious fears of inadequacy and incompetence, which serve to undermine a group's ability to perform rationally. While not prevalent in modern organizational thinking, training in Bion's theory is still available through the Tavistock Institute of London and at the A. K. Rice Institute in the United States.

John Hemphill's Situational Factors in Leadership.

Hemphill's historic (and classic) 1949 manuscript, *Situational Factors in Leadership,* outlines his research on the identification, measurement, and evaluation of 15 group dimensions for the description and differentiation of numerous types of groups (see Table 1). This work was a marked departure in the study of groups and was based on considerable conceptual, empirical, and, at that time, innovative methodological applications utilizing large samples and stringent methods of evaluation. Among the many groups described in his work, Hemphill (1949) attributed the differences in the 15 dimensions to variation in leadership behavior. The contribution of Hemphill's work lies in the illustration of groups with different functions in a variety of organizational settings, with groups profiled according to the 15 dimensions—many of which are still useful and the subject of research today.

Some lessons learned:

Lessons Learned 7: Collectives are a source of sustained competitive advantage (Scanlon; Frost, 1974).

Lesson Learned 8: Work groups can be studied as open, sociotechnical systems (Trist e al., 1997).

Lesson Learned 9: To succeed, collectives must freely exchange resources across team members, while maintaining their interpersonal relationships. This lesson is even more salient when teamwork is made increasingly complex by the introduction of technology or additional tasks (Tavistock coal studies; Trist et al., 1997).

Lesson Learned 10: Groups and teams are not the domain of any single discipline but rather should be the focus of an integrated multidisciplinary effort (Bion, 1948).

Lesson Learned 11: Leadership behaviors matter in groups (Hemphill, 1949).

A new momentum in team and group research: 1950's – 1960's

The Human Resource Paradigm & Teams

In the mid twentieth century momentum was building for the humanistic ideals espoused by a growing community of researchers (e.g., Follett, Roethlisberger, McGregor, Janis). At the heart of this movement was a historically unprecedented assumption that organizations exist to serve human needs (Shafritz & Ott, 2001). This new line of thinking served to change the nature of the relationship between employers and employees from one of dependence to codependence. However, while the set of assumptions underlying the human resource / organizational behavior (OB) perspective stood in stark contrast to the assumptions underlying the classical school's assertions, the new mantra of codependence only served to reinforce the importance of the two enduring themes of coordination and cooperation previously advanced by the classical thinkers.

The distinguished scientists noted here, along with hundreds of others, created a wave of research activity that produced an enormous body of supporting literature and methods. Slowly, these tools, techniques, strategies, and methods began to permeate the practices of many organizations. One manifestation of this movement is witnessed in the wide range of motivation theories that served to shape how organizations went about proactively influencing employee attitudes, perceptions, beliefs, and motivations (e.g., Maslow's need hierarchy [Maslow, 1943], Herzberg's [Herzberg, 1966] two-factor theory). Also, these initial theories such as Herzberg's two-factor theory served as a platform for future endeavors (see Hackman & Oldham's [1975] job characteristics theory) that would eventually illuminate the complexities of motivated behavior in groups and teams.

One example of how the OB paradigm is intertwined with teamwork is found in participatory work practices. Participatory work practices encourage the use of teams because collectives can

draw from their diverse reservoir of resources to better accept responsibilities traditionally reserved for management. For organizations, this meant that layers of middle management could be unceremoniously stripped away as empowered teams began to delegate work, determine strategies for task accomplishment, and resolve interpersonal conflicts. Certainly, this new state of affairs made for leaner, more adaptive organizations. In fact, recent research suggests that including employees in decision-making results in higher levels of cooperation, satisfaction, productivity, and quality and lower levels of absenteeism and turnover (Manz & Neck, 1995).

Key Contributors

Douglas McGregor. If Follett and the members of the Mayo team at Hawthorne can be considered the initial visionaries of the OB movement, then the defining statement of this paradigm was advanced by Douglas Murray McGregor in 1957. McGregor's Theory X and Theory Y radically reframed thinking about the nature of workers and what they could be empowered to accomplish. McGregor (1957) proposed that the managerial assumptions about the nature of workers comprising both Theory X (i.e., employees are unmotivated and require constant direction) and Theory Y (i.e., employees are active participants that can be developed and trusted) would, over time, crystallize to form self-fulfilling prophecies and thereby effect worker behavior.

While Theory X owed its epistemological roots to classical organizational theory, Theory Y was a revolution that suggested workers could be trusted to provide their own self-direction, would seek out responsibility, and should develop and grow while achieving financial objectives. These basic tenets became the bedrock upon which future performance arrangements (e.g., self-managing work teams, distributed teams, shared team leadership) would be built.

Irving Janis. If McGregor's assertions were the flagship for the promise of OB, Janis' ideas were equally powerful in warning about the dangers of interdependent interaction. Although the treatment of Janis (1972) in this chapter technically belongs in the next period of time reviewed in this chapter (i.e., 1970s – 1980s) his research on social conformity in groups so contrasts the optimism exuded by McGregor's research that we felt a discussion of these issues within this subsection offered the reader an appropriate amount of historical balance.

Janis' (1972) research addressed 'groupthink', described as the train of thought that highly cohesive group members adopt when concurrence seeking takes precedence over reality. This line of research suggests that decision making in cohesive groups during crisis situations (e.g., Vietnam, Cuba Invasion, Columbia Shuttle) can create a pressure to conform at all costs. Furthermore, once a decision is collectively reached, entities suffering from groupthink act to suppress dissent, overestimate their probability of success, and establish mindguards. Fortunately, Janis suggests that the groupthink phenomenon is relatively easy to identify and can be effectively dealt with (see Janis, 1972).

Group Interventions

T-Groups. Training groups (T-groups) also became prominent in the United States post war era. In 1947, the National Training Laboratories (NTL) developed a new training innovation known as T-groups. T-groups were an early precipitation of team building. Based on theories of human behavior, T-groups facilitated interaction by enlightening members about options for their behavior in groups. In other words, group members were given training on how to look within themselves and their groups in an effort to gain a deeper level of insight and thereby appreciation. Core components of this process included self-observation, self-disclosure, vicarious learning, and catharsis (Rogers, 1970).

Over time, T-groups led to the use of sensitivity groups, implemented to explore and enhance group members' interpersonal sensitivity. T-groups also gave way to team building which overcame some of the problems inherent to the process (Cranton, 1996). While T-groups are still in use today, the 1970's brought about a backlash against touchy-feely training interventions that served to notably curtail the formerly wide spread use of this approach (see Highhouse, 2002b).

Quality Circles. Quality Circles (QCs) are another intervention with a long and storied history. The origination of QCs is often attributed to the U.S. in the 1950's. However, medieval European guilds also adopted a QC approach in order to guarantee the quality of their crafts (McCarthy, 2002). The modern form of Quality circles (QCs), also known as quality control circles (QCCs) or process improvement groups (PIGs), emerged from Japan's total quality control movement in the 1960's (Glassop, 2002). Today, QCs are most closely associated with Professor Kaoru Ishikawa, who became known as the pioneer of QCs after this approach was re-exported to the West in the 1960's. QCs were developed as a training mechanism for promoting departmental quality. The key to this approach is the use of statistical processes to eliminate problems caused by unwanted variations in the work system.

In the 1960's, QCs were conducted outside normal work hours, but have since been intertwined with ongoing operations in order to better address a wider variety of problems in a shorter period of time. Typically, modern QCs are comprised of small groups of co-located workers who meet for a short amount of time each week to identify, analyze, and solve various work-related problems by making recommendations to management (Tang & Butler, 1997). By the mid 1980's, QCs had been Americanized and were one of the most promising team related strategies implemented in industry (see Bettenhausen, 1991).

Delineating the Nature of Teams

In the mid 20th century, team theory building, research, and practice began to accelerate, albeit slowly at first. The 1950s and 1960s witnessed the proposal of the initial distinctions between groups and teams and team characteristics and task characteristics by team researchers. An important example is the research of Roby and Lanzetta (1957) who studied the negative impact of task overload on team communication and performance. Also, Briggs and Johnston (1967) examined the importance of team member backup behavior in reducing flight errors. This line of research is especially important to the history of teams, given that today it is generally believed that effectiveness criteria are contingent upon the nature of the team under consideration (Cohen & Bailey, 1997).

A major organizational development during this period was the establishment of the Crew Consortium Research Laboratory, as a unit of the Air Force Personnel Research and Training Center in San Antonio, Texas. This was a primary effort to conduct experimental studies of group formation and functionary, team performance, and team effectiveness. The staff included Robert French (Director), Thornton Roby, John Lanzetta, Donald Forgays, Lemar Berkowitz, Bernard Lery, Seymour Rosenberg, and others. The research of Roby and Lanzetta (1957) provides only one example of many, developed in this laboratory and published in scientific journals. The laboratory provided one of the early opportunities to study the formation, assembly, simulation, and performance of real teams in relevant operational settings. These individuals moved on to distinguished academic careers.

In time, the research of Roby and Lanzetta (1957) and Briggs and Johnston (1967) would serve as a platform for researchers attempting to build a science of teams. For example, distinctions would be made between taskwork and teamwork. Specifically, taskwork refers to the

individual behaviors team members engage in while enacting their skills during team performance. Taskwork is “a team's interactions with tasks, tools, machines and systems" (Bowers, Braun, & Morgan, 1997, p. 90). Teamwork behaviors include backup behavior, coordination, collaborative problem solving, and dynamic reallocation of function. Moreover, the behaviors that comprise teamwork can have an individual or team level referent. The findings from this research have been essential in designing, delivering, and evaluating team processes, structures, and training interventions.

Framing Group & Team Effectiveness

In the previous section, we noted how researchers began to address the nature of teams in the mid-twentieth century. As findings accumulated from these programmatic initiatives it became apparent that a framework was necessary to house the myriad of constructs comprising and impinging upon team effectiveness. Researchers such as McGrath (1964) analyzed group interactions through organizing frameworks in an effort to illuminate those illusive factors that make some groups effective (i.e., process gains), while others experienced process losses and failure. From these efforts, the input, process, output (IPO) model was advanced as a means of framing group and team effectiveness (McGrath, 1964).

An IPO approach to understanding team effectiveness shed new light on the moment to moment interteam and intrateam processes that comprise teamwork. While the idea seems simple, it was revolutionary for those concerned with teamwork to have moved beyond the basic manipulation of inputs as a means of controlling outputs and towards understanding the ‘black box’ of team process. This movement is reflected today in the growing number of conceptualizations of team performance and effectiveness which show an appreciation for the importance of time and its relevance for dynamic processes and their interrelationships

(e.g., Dickinson & McIntyre, 1997; Marks et al., 2001; Shiflett, Eisner, Price, & Schemmer, 1985). Specifically, an IPO approach is flexible enough to model the more dynamic aspects of effectiveness (e.g., feedback loops, recursive processes, linear time, reciprocal linkages).

The use of an organizing IPO framework was indeed a major advancement in the group and team domains (e.g., Ilgen, Hollenbeck, Johnson, & Jundt, 2005). In fact, the importance of the IPO model to the study of group effectiveness can not be understated as this approach is widely utilized today as an organizing framework. Over time, findings have supported the basic model, which outlines how team performance is affected by system inputs (e.g., individual, group, environmental) acting through interaction processes. However, much debate remains about the nature of the construct relationships modeled within an IPO approach. This particular issue concerns whether processes are best conceptualized as mediators or moderators in the IPO chain (Kozlowski & Bell, 2002).

Some lessons learned:

Lesson Learned 12: Groups can help foster quality products and processes in organizations.

Lesson Learned 13: It is essential to recognize the importance of actively promoting group interaction and collaboration (Cranton, 1996; Briggs & Johnston, 1967).

Lesson Learned 14: Work groups are an intuitive unit that can address and solve complex organizational problems (Tang & Butler, 1997).

Lesson Learned 15: Not all teams are created equal; they vary in the degree of task interdependence.

Lessons Learned 16: Group and team effectiveness can be framed as a series of inputs, processes, and outputs (IPO) (McGrath, 1964).

Lesson Learned 17: Framing group and team effectiveness from an IPO perspective facilitates modeling the dynamic, 'black box' of team process (McGrath, 1964).

The influence of technology on teams and groups: The 1970's-1980's

Technological Advancements

In the late 1940s, the first computer was designed and just a decade later the launch of Sputnik started a space race. However, it would take until the 1970's for technology to fully captivate the public's attention. Throughout the 1970's, technological innovations were made daily, spread widely, and began to affect our everyday lives. Video games, optical fiber, barcodes, the fleet of space shuttles, genetic engineering, and the artificial heart, are just a few examples of how technology began to change the world. Once again, a revolution had put teams back in the spotlight and complex context had ensured they were here to stay.

A building wave of team research began to mount in part because from designers to end users teams were intertwined with technology. Technological advancements changed the way (1) jobs were performed, (2) people communicated, and (3) made decisions. Advances in technology opened the door to address previously unapproachable problems (e.g., space travel, genetic sequencing, ocean floor mapping). However, the complex nature of these challenges required more than innovative technological solutions. Problems of this magnitude required teamwork. Furthermore, society's hopes rested on the shoulders of expert teams, not just teams of experts.

During this historical period, the world was becoming increasingly characterized by a freer and liberated spirit and this change in attitudes inevitably influenced organizations to adopt humanistic values. One manifestation of the changing perspectives is found in the widely popularized belief that employees are not just interchangeable cogs in the organizational wheel

(Weber, 1972). This belief reflects the humanistic ideals that were underlying a majority of the work conducted in the previous decades, as well as much of the research in vogue at the time.

The resurgence in team research also led to much theoretical and applied advancement. A wide range of research was undertaken to investigate everything from sociopsychological constructs to the deeper cognitions that permeate information processing in groups. The following subsections illustrate a representative sample of this theoretical work and illustrate applied examples for the reader in an effort to demonstrate the successful application of theory to industry.

Work Groups in Plants

General Foods. An explosion of team implementations took place in the 1970's in response to new technologies utilized in manufacturing plants and warehouses. These technological innovations resulted in more complex workplaces and necessitated the need for teams to handle the operational complexity and to maximize the value of these new technologies. For example, the General Foods plant in Topeka, Kansas is one of the most widely cited case studies of the implementation of a humanistic management approach. Specifically, "the plant has taken a position in many minds as a classic demonstration of the power of a system-wide organizational innovation, quality-of-work-life interventions, and sociotechnical systems approaches" (Whitsett & Yorks, 1983, p. 93).

The planners of the Topeka plant aimed to create an organization that would not only be efficient, but would also provide employees with jobs that would be rewarding, interesting, and challenging. Their most innovative vision was to have the plant evolve to a self-managing multi-team system. This vision, once taken to fruition, would largely alleviate the need for formal management, give all personnel an equal voice in the plant activities, and tear down barriers by

removing differential status symbols common in most plants (i.e., reserved parking places). Upon implementation, this vision was known as the Topeka System.

The Topeka System consisted of several components addressing the work design, supervisory structure, social system, and reward system. The work design component of this process included the use of three types of teams that covered the entire phase of the plant's operation: processing, packaging/shipping, and office support. Each team, comprised of approximately 14 members, was empowered to control how tasks would be completed. The teams also managed the process of interviewing new job applicants, redistributing work when members were absent, scheduling breaks, establishing and changing schedules, conducting performance evaluations, and counseling those not meeting established quality control standards. The supervisory structure under the Topeka system was also novel in that team leaders were chosen to be "managers of people rather than technical specialists" (Whitsett, & Yorks, 1983, p. 96).

The reward system is unquestionably the most innovative aspect of the Topeka system because it was designed to compensate employees for mastering both individual and team KSAs. Specifically, the compensation system was comprised of four pay rates that varied on a sliding scale ascending as a function of employee knowledge and skill development. This compensation system was designed to encourage employees to proactively engage in self-development rather than reactively waiting for formally sponsored organizational training to be administered.

Each employee in the Topeka System started at the same rate. An employee could then be raised to the "single job rate" when it was determined by the team leader that they had mastered the initial job assignment. The next rate was the "team rate" and was earned when the employee mastered the jobs of all the team members. Finally, a "plant rate" could be earned when an

employee learned all of the jobs in the entire plant. There was also an “add-on” rate if the employee had some special capability, such as electrical maintenance (Whistlett & Yorks, 1983).

Although the Topeka plant was intended to be innovative and was eventually modeled in other General Foods plants, the dream was never fully achieved as it had been envisioned (Whitsett, & Yorks, 1983). However, it should be noted that team members in the Topeka System responded enthusiastically to this intervention. In fact, product quality was uniformly high (e.g., 10 to 40 percent greater than comparable organizations), absenteeism and turnover were significantly below industry standards, and theft was reported to be negligible (Management Review, 1972).

Volvo. Volvo has a long history of employee-centered innovation that is now in part credited for its status as a world-class automaker and organization. Pehr Gyllenhammer became Volvo’s CEO in 1971 and believed that if Volvo was to remain a world-class competitor, the corporate culture had to change (Bernstein, 1988). Under Gyllenhammer’s leadership, Volvo attempted to improve job satisfaction through job rotation, management-employee councils, small work groups, modern technology, and employee-oriented facilities (Gyllenhammer, 1973).

One example of Volvo’s innovativeness took place in the upholstery division at the Torslanda plant (Bernstein, 1988). Incumbents of this division often complained of work related muscle strains due to the repetitious nature of their work. In order to remedy this problem, the workers began experimenting with groups and found that through job rotation and cross training, their muscle strains were largely eliminated. Furthermore, the implemented work changes resulted in secondary benefits such as improved group cohesion and a renewed emphasis on quality. The Lundby plant is a second example of teams implemented in Volvo (Bernstein, 1988). Managers of the Lundby plant empowered teams to be responsible for individual

functions or parts of the car. The implementation of these 'functional teams' resulted in higher productivity and quality by increasing employee involvement.

A third example of how Volvo utilized teams to achieve its organizational objectives is found in the Kalmar plant, built in 1974. The Kalmar plant relied heavily on teams. However, unlike the General Foods Topeka plant discussed earlier, this facility was specifically designed to facilitate teamwork. For example, the physical design of the plant encouraged greater employee productivity by providing work team collaboration areas, more windows and sunlight, and a less pressured work environment. Furthermore, the work itself was organized so that each group was responsible for a particular portion of the car. This method of automobile assembly provided the group with a sense of task identity (see Hackman & Oldham, 1975) via a product which they could craft from start to finish. Today, Kalmar exemplifies a successful team implementation that was due in large part to organizational support for teams, joint decision-making, communication, flexible work schedules, and new production technologies (Bernstein, 1988).

Some lessons learned

Lesson Learned 18: Manage the people in teams and groups, not just the technology (Whitsett & Yorks, 1983).

Lesson Learned 19: Organizations which use teams can be efficient while providing employees with jobs that are rewarding, interesting, and challenging (General Foods).

Lesson Learned 20: Simultaneously change many aspects (e.g., work design, supervisory structure, social system, rewards) of the same system to foster teamwork (Topeka System).

Lesson Learned 21: Compensate employees for mastering both individual and team-based KSAs (Topeka System).

Lesson Learned 22: Functional teamwork (e.g., empowering teams to be responsible for individual functions) increases involvement (Hackman & Lawler, 1971)

Lesson Learned 23: Teams share the load because team members can compensate for one another to ensure everyone remains healthy and productive (Bernstein, 1988).

Lesson Learned 24: Job/task characteristics can be manipulated to induce desired cognitive and affective states and thereby increase desired individual, as well as team, outcomes (Hackman & Oldham, 1975).

The booming and explosive 1990's & beyond

Team Maturation

The systematic study of group and team maturation has been ongoing for almost four decades. One of the first researchers to engage in this line of work was Tuckman (1965), who examined the maturation of T groups, therapy groups, and laboratory groups. From his research findings Tuckman delineated and advanced a stage model of group development. This classic model illustrates how groups develop over time by transcending from one stage to the next (see Wheelan, 1994). Four stages are typically proposed in most maturation conceptualizations including forming, storming, norming, and performing. Occasionally, a fifth stage (i.e., adjourning) is also utilized to describe the process via which groups dissolve.

While Tuckman's seminal work has served as a platform for current research initiatives, its primary focus on interpersonal conflict has served to limit its wide spread applicability to understanding teamwork. As a result of this restricted focus, a stage approach usually fails to model many of the important aspects of context, tasks, and the often discontinuous configural compilation process via which team member inputs, dyadic role exchanges, and team level contributions emerge to form team performance (Kozlowski & Bell, 2002).

Fortunately, research is underway to examine how teams form, evolve, and perform in naturalistic settings (see Ancona & Chong, 1999; Edmondson, 1999; Gersick, 1989; Harrison, Mohammed, McGrath, Florey, & Vanderstoep, 2003; Ilgen, 1999; Kozlowski, Gully, Nason, & Smith, 1999; Morgan, Salas, & Glickman, 1993; Sheard & Kakabadse, 2002). This research has been undertaken to address some of the above noted weaknesses of the classic stage models by approaching maturation and performance from a more integrative, contextualized perspective.

Some of the findings from this new breed of research suggest team development may be driven as much by temporal entrainments as by interpersonal dynamics (Gersick, 1989). Furthermore, theoretical and empirical research suggests that maturation requires the development of both taskwork and teamwork KSAs (Glickman, Zimmer, Montero, Guerette, Campbell, Morgan, & Salas, 1987; Marks et al., 2001; Morgan et al., 1993). In addition, Kozlowski and colleagues (1999) advanced a team compilation model that includes aspects of both team development and performance.

Team Performance & Effectiveness

All of the research discussed in this historical review contributes to the collective understanding about team performance and stakeholder judgments about the effectiveness of team performance. However, a growing number of initiatives have directly addressed the complexities of team performance and effectiveness by advancing integrative theories on the issue (see Dickinson & McIntyre, 1997; Gersick, 1988; Gladstein, 1984; Hackman, 1987; Marks et al., 2001; Nieva, Fleishman, & Reick, 1978; Tannenbaum, Beard, & Salas, 1992). One comprehensive synthesis of this literature identified over 800 empirical studies addressing some aspect of teamwork and 138 models and frameworks of team performance and effectiveness that have been advanced in just the last 25 years (Salas, Stagl, Burke, & Goodwin, in press). In fact,

the results of this review suggest there was a 257% increase in the number of models and frameworks of team performance and effectiveness advanced in the 1990s as compared to the prior decade.

Although even a surface level discussion of these 138 models and frameworks of team performance and effectiveness are beyond the scope of this chapter, one comprehensive initiative undertaken by Campion and colleagues is briefly reviewed as an illustrative example of this body of research. Campion et al. (1993) advanced a 'meta' model that highlights the importance of five categories of variables (i.e., job design, interdependence, composition, context, process) which are proposed to directly impact team effectiveness. Each of these five categories includes a number of variables. For example, the job design category includes five specific variables: self-management, participation, task variety, task significance, and task identity. One of the greatest strengths of Campion and colleagues 'meta' model is its empirical support accumulated from two diverse samples of teams (see Campion et al., 1993; Campion, Papper, & Medsker, 1996).

In addition to the above noted models and frameworks which attend to the nature of routine performance, a growing but sporadic body of theory and research has also highlighted the importance of adaptability, adaptive team performance, and team adaptation to fostering team effectiveness (Arrow, McGrath, & Berdahl, 2000; Chen & Ployhart, 2004; DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004; Entin & Serfaty, 1999; Klein & Pierce, 2001; Kozlowski, Gully, Brown, Salas, Smith, & Nason, 2001; LePine, 2003; Marks, Zaccaro, & Mathieu, 2000; Pulakos, Arad, Donovan, & Plamondon, 2000; Pulakos, Schmitt, Dorsey, Arad, Hedge, & Borman, 2002; Waller, 1999). Despite the importance of team adaptation to team effectiveness, little is currently known about the nature of this construct or the proximal and distal constructs contributing to it (Campbell & Kuncel, 2001). In fact, to date team adaptation has not been the

subject of systematic investigation in the behavioral sciences. Thus, this is one area which is ripe for theory building, laboratory research, and qualitative and quantitative field studies.

Team Types & Taxonomies

The current science of teams is the result of a century of effort by scientist practitioners from a wide range of disciplines and philosophies engaging in theory building, research, and practice. Unfortunately, however, the varying assumptions, approaches and interpretations of these theoreticians, researchers, and practitioners have in part served to create a diverse and sometimes unconnected body of team research findings. In an effort to impose structure on this diffuse body of knowledge, researchers have undertaken taxonomic initiatives (e.g., Fleishman & Zaccaro, 1992). The purpose of these endeavors is to define and classify elements in order to infer general statements about property classes and class interrelationships (Fleishman & Quaintance, 1984). Taxonomic initiatives are increasingly important for capturing and codifying accumulated findings because there is an accelerating storm of team research and application as we begin the new millennium. In fact, encoding what is currently known is a challenge for team researchers, as well as, the behavioral sciences in general (Campbell, 1990).

Fleishman (1975) and Fleishman and Quaintance (1984) described an extensive conceptual and empirical process to develop and assess taxonomies of human performance that would enhance predictions and generalizations about human performance. Primary considerations were the adoption of task definitions that permitted reliable task descriptions, the inclusion of categories that were neither specific nor too general, the development of measurement systems that correspond to taxonomic categories, and the establishment of criteria for psychometric assessment of these approaches.

Fleishman and his colleagues extended this work on individual performance taxonomies to the analysis and assessment of team performance (Nieva et al., 1978; Shiflett et al., 1985; Cooper, Shiflett, Korotkin, & Fleishman, 1984; Fleishman & Zaccaro, 1992). This multi-year effort, initially supported by the Army Research Institute and the Air Force Systems Command, involved considerable conceptual and methodological development and application of a measurement system to evaluate interactions among team descriptions, their reliability and independence, and their validity for predicting team performance (e.g. using Army construction units, command & control teams, etc.). It appears that a meaningful taxonomy and measurement system for describing and evaluating team performance functions, emphasizing requirements for coordinated actions of team members to produce synchronized output, can be developed. Table 2 provides a recent outline of this taxonomy (Fleishman & Zaccaro, 1992).

One area that has particularly benefited from the application of taxonomies is the study of team types. Over the last quarter of the twentieth century, organizations have begun to increasingly utilize teams for many different functions. In response, a building wave of research has been undertaken and theoretical advancements have been made by directing attention toward different "types" of teams (e.g., action teams, self-managing teams, management teams). In fact, over the past century, dozens of team types have been advanced.

In an effort to impose structure on the multiple labels that exist in the current literature base, several researchers have created team typologies (e.g., Argote & McGrath, 1993; Cohen & Bailey, 1997; Devine, 2002; Devine, Clayton, Philips, Dunford, & Melner, 1999; Hackman, 1990; Mohrman, Cohen, & Mohrman, 1995; Sundstrom, DeMeuse, & Futrell, 1990; Sundstrom et al., 2000). For example, one widely accepted team typology was presented by Sundstrom and colleagues (2000) who condensed a myriad of team labels into 6 main types of teams:

management, project, product, service, action, and parallel. These researchers chose to classify and describe team types on the basis of differentiation, external integration, and work cycles (see Sundstrom et al., 2000). A similar initiative by Devine (2002) resulted in 14 team types. Devine's (2002) effort emphasizes teams performing day to day operations, teams engaged in intellectual work, and relevant contextual forces.

It is important to note that all of these types of teams serve different purposes. Furthermore, the organizational culture, team size and structure, organizational support, to name a few, impact the success of each type of team within their own organization. This is critical because team effectiveness criteria vary with the team type under consideration (Cohen & Bailey, 1997). Thus, the efforts just reviewed and other similar initiatives are essential to understanding team effectiveness and to designing interventions to foster teamwork.

In addition to the general team typologies reviewed above, a number of specific types of teams have been investigated. One of the most important types of teams operating in dynamic environments is the self-managing work team (SMWT; Pearce & Ravlin, 1987). In response to modern, chaotic context, organizations are enabling and empowering team members to make real time tactical decisions. A primary reason for the growing base of support behind this approach is that organizations are increasingly cognizant of the value gained from enabling team members to act in a proactive, adaptive manner to capitalize on opportunities while simultaneously reducing the losses incurred when it is necessary to react to events.

Self-managing work teams are empowered to manage themselves and therefore continually focus on improving their own work processes. As such, these teams are responsible for much of what would traditionally be accomplished by hierarchical leadership outside the team. Although there is some degree of variability, typical characteristics of SMWT include: (1) a distinct set

tasks that workers can identify with, (2) members with a variety of skills that are related to the task, and (3) high levels of empowerment (i.e., discretion over how work is done) (Hackman, 1987). Others have argued that these teams: (1) tend to have a flat structure, (2) shared leadership, and (3) emphasize leader coaching and facilitation (Wellins, Byham, & Wilson, 1991). Although SMWT have high levels of empowerment, they are still subject to some external control. For example, self-managing work teams do not typically set strategic objectives or compensation policies (Stewart, Manz, & Sims, 1999).

We suspect that the importance of self-managing work teams to organizations will continue to grow in step with the accelerating rate of operational complexity. In fact, self-managing work teams are ubiquitous in large organizations as witnessed by a recent report indicating that 47% of Fortune 1000 companies are utilizing them (Cohen, Ledford, & Speitzer, 1996). It seems that despite arguments to the contrary (see Mullen & Cooper, 1994) organizations are actually 'In Search of Excellence' (see Peters & Waterman, 1982). A major concern for such systems is the development of their members' capacities so that they can fully capitalize upon their empowerment; this and other training issues will be addressed in the next section.

Hierarchical Team Decision Making

Decision-making is ubiquitous to teamwork and is one of the most frequently studied phenomena in the behavioral sciences. Unfortunately, however, the vast majority of research to date has applied the social decision scheme model and the social transition scheme modelto understanding the consensus decisions of small groups (e.g., jury decisions) (see Davis, 1992). Collectives such as juries can be characterized as a collection of largely independent members, who are by nature equal in status and expertise. In contrast, teams are typically hierarchically structured, reflecting important differences in member status and expertise. Moreover, teams

have much higher levels of goal, task, and feedback interdependencies than do juries and other groups. Therefore, the generalizability of the inferences which can be drawn from the accumulated findings of studies investigating consensus decisions in groups, to more highly interdependent collectives such as teams, remains an unanswered empirical question.

Fortunately, researchers have begun to move beyond the study of group decisions to systematically investigate and model the decisions of teams (see Ilgen, Major, Hollenbeck, & Sego, 1991). For example, recent research has advanced and tested multilevel theories of team decision making (see Hollenbeck, Ilgen, LePine, Colquitt, & Hedlund, 1998; Hollenbeck, Ilgen, Sego, Hedlund, Major, & Phillips, 1995). Hollenbeck, Ilgen, and colleagues theory of multilevel team decision-making extends prior theories of group decision-making by: (1) emphasizing the inputs into a decision rather than beginning with decision outcomes, (2) characterizing decision alternatives along a continuum rather than a dichotomy, and (3) modeling both individual team member and team level decision processes. This seems to be a particularly promising approach to understanding and improving team decision-making, as witnessed by research results suggesting operationalizations of the constructs core to this theory can account for over 50% of the variance in team performance.

Team Training & Development

Team Training. Although a body of literature addressing training emerged in the 1920s (Ford, 1997), team training has just begun evolving into an integrated science during the past two decades (see Campbell & Kuncel, 2001; Kraiger, 2002; Salas & Cannon-Bowers, 2001; Tannenbaum & Yukl, 1992). One area that has received considerable attention concerns the design, development, and delivery of team-based instructional strategies. Throughout the late 1990's and into the 21st century, a number of strategies for team training began to be applied that

emphasized teamwork competencies in aviation, healthcare, the military, and the nuclear power industry, to name a few. Team-based instructional strategies are forged from a set of methods, tools, team-based KSAs, and targeted learning outcomes. These components, when considered with organizational factors (e.g., resources, policy, size), help determine which strategy is appropriate for a given team or work group.

Another important development was the U.S. Navy's Tactical Decision Making Under Stress (TADMUS) project which began in 1990 as an interdisciplinary research initiative undertaken to understand decision-making under stress and provide instructional strategies for teams. It was the Navy's expectation that adopting a systematic approach to illuminating the complexities of decision making in high pressure environments would lead to theoretical breakthroughs that could be translated into practical products for future use in team training and human factors technology. In retrospect, they were more than correct, as the science of team training has markedly benefited from the broad range of strategies developed through TADMUS research (see Cannon-Bowers & Salas, 1998).

The multiple, team-focused strategies available today are historically unprecedented. Organizations now have tools, techniques, and strategies that can train both team members and entire teams beyond simple task proficiency and towards skill generalization and adaptation. A particular exemplar, successful in the aviation industry and more recently in healthcare, will be discussed in greater detail below.

Crew Resource Management. A major movement in team research took place in the skies. Cockpit or Crew Resource Management (CRM) originated from a NASA workshop, *Resource Management on the Flightdeck,* in 1979 (Helmreich, Merritt, & Wilhelm, 1999). The program was held in response to the findings that a majority of aviation crashes could be linked to human

error and specifically to failures of interpersonal communications, decision making, and leadership (Helmreich et al., 1999; Salas, Wilson, Burke, & Bowers, 2002). In fact, reports suggest that 60% to 80% of aviation crashes can be traced to pilot error, many of which are due to a lack of coordination in the cockpit (FAA, 1998). The FAA has proactively addressed this problem with CRM training as "a way to prevent aviation accidents by improving crew performance through better crew coordination" (FAA, 1998, p. 4). CRM was developed to improve and enhance the use of all resources available on the flightdeck, which includes teams, systems, interpersonal aspects of flight operations, information between team members, and information exchanged between systems and team members.

CRM has contributed to the science of team training by providing a systematic approach to optimizing team performance. In fact, CRM can be utilized to develop backup behavior, feedback, performance monitoring, and coordination, all processes now accepted as necessary for effective team performance. CRM remains a dynamic and encompassing training program for team effectiveness in the cockpit and can be beneficial to a variety of industries, such as natural resource refinement on oilrigs (Flin, 1995) and health care (Gaba, Howard, & Fish, 1994). For more organizational applications of CRM see recent publications by Salas, Bowers, and Edens (2001) and Merritt and Helmreich (1995).

Team Building. Team building interventions were extremely popular in the 1980's but have received relatively less attention recently. Perhaps this is because the economic prosperity of the 'me generation' of the 1980's led to corporate retreats and other 'touchy-feely' interventions similar to T-groups. By nature, team building stresses the importance of an action research model of data collection, action planning, and feedback. While a great deal of research on team building was conducted in the 1960's through the 1980's, more recent endeavors are characterized by a

concentration on increasing role clarification and social aspects of teams. While team building has been found to lead to many positive affective outcomes for team members, questions remain about the behavioral changes of teams following these interventions (Salas, Rozell, Mullen, & Driskell, 1999).

Team Performance Measurement

As teams began to be better understood through the 1990's, the complexity of teamwork led to a growing concern for how to measure team process and performance (see Brannick, Salas, & Prince, 1997; Day & Lance, 2004). By this time, researchers and practitioners had come to the realization that teams are different from both their constituent members and from each other. Previously, highly informal and thereby subjective measures were utilized to evaluate team effectiveness such as asking members for their impressions of the team's performance or by observing the quality or frequency of specific behaviors (Weingart, 1997).

Overtime, however, it became apparent that observable behaviors are inadequate to understand team performance. A team's knowledge and attitudes are also important, requiring new and more formalized measurement techniques. Subsequently, there have been advancements in measurement ranging from event-based measurement (e.g., systematically creating opportunities to observe behaviors of interest), real-time assessments (e.g., automated data collection of team members inputs such as key strokes), and improvements in observation and coding schemes (e.g., use of behavioral rating scales). Researchers also learned that team performance measurement must be multilevel and dynamic in order to capture teamwork (Cannon-Bowers & Salas, 1997; Klein & Kozlowski, 2000). This suggests that not only must the performance of the team be evaluated, but also the performance of every individual on the team (Tesluk, Mathieu, Zaccaro, & Marks, 1997). Measures must also be flexible enough to capture

the dynamic, recursive nature of teamwork. However, even as performance measures have become the focus of research, there is still a considerable gap between what researchers have found and what practitioners typically utilize in their daily affairs.

Shared Cognition

Recent team studies are increasingly incorporating constructs, theory, and findings from cognitive psychology (e.g., transactive memory, shared mental models, metacognition) (Austin, 2003; Griffith & Neale, 2001; Lewis, 2003; Moreland, 1999; Salas & Fiore, 2004; Wegner, 1987). The emphasis on the K (i.e., knowledge) in teamwork KSAs is indeed as important as teamwork skills. From this movement, new tools for knowledge elicitation and the measurement of knowledge structures (e.g., cognitive mapping, card sorts) have been advanced. These breakthroughs have served to make the study of team cognition accessible to a wider audience.

Many researchers beginning in the 1980's and continuing into the 21st century began to examine shared cognition, which refers to overlapping, accurate, flexible knowledge structures (e.g., Campbell & Kuncel, 2001; Cannon-Bowers, Salas, & Converse, 1993; Ensley & Pearce, 2001; Entin & Serfaty, 1999; Hinsz, Tindale, & Vollrath, 1997; Klimoski & Mohammad, 1994; Orasanu, 1990). Research in the late 1980's referred to constructs such as "team mind" (Klein, 1998) or "transactive memory" (Wegner, 1987). Early empirical work found support for the role of shared cognition in a team's ability to coordinate implicitly (Kleinman & Serfaty, 1989). Further evidence reinforced this idea, stressing the benefit of shared cognition in predicting what team members will do, given unexpected or novel events (Orasanu, 1990).

As research on shared cognition progressed, it came to be largely focused on shared mental models (SMM). Shared mental models are defined as "organized knowledge that is shared by team members" (Orasanu & Salas, 1993, p. 8). However, Cannon-Bowers and colleagues (1993)

expanded that definition to specify that “in addition to shared knowledge team members must hold shared expectations that are generated from this knowledge” (p. 228). This line of research is founded on the premise that team synchronicity (i.e., process gains) can be achieved when team members develop clusters of shared and accurate instantiated knowledge structures (i.e., shared mental models). Essentially, team members operating from a shared mental model have a common conceptual framework which enables them perceive, interpret and respond to dynamic environments in a coordinated, adaptive fashion.

It has been repeatedly suggested that SMMs can serve to ameliorate the negative effects of time pressure and stress by facilitating implicit coordination and communication (Kleinman & Serfaty, 1989). In fact, there is an accumulating body of evidence suggesting the sharedness and accuracy of SMMs serve to facilitate team performance and effectiveness (Mathieu, Heffner, & Goodwin, 1996; Marks, Sabella, Burke, & Zaccaro, 2002; Stout, Cannon-Bowers, Salas, & Milanovich, 1999).

Distributed & Virtual Teams

Globalization and technological innovations have enabled organizations to seek out the best talent regardless of where it is located and organize that human capital via distributed performance arrangements (e.g., distributed teams, virtual teams). Virtual teams are defined as teams where “members' primary interaction is through some combination of electronic communication systems, such as telephone, fax machine, e-mail, and computer-based video conferencing” (Townsend, DeMarie, & Hendrickson, 1996. p. 122) in which team members “are mediated by time, distance, or technology” (Driskell, Radtke, & Salas, 2003, p. 3). Not surprisingly, organizations are implementing these types of teams without regard for the science behind their effectiveness. Despite their increased use, we do not have a good understanding of

the impact that becoming ‘virtual’ has on team member and team processes and performance. In fact, most of what is currently known has been generalized from research with collocated teams.

Given what is known about collocated teams and their coordination requirements, it has been argued that teams separated by time and or space have additional demands placed on them during distributed interaction. For example, team researchers have explored the absence of cues (e.g., facial expressions) for distributed teams. The lack of stimuli input taxes the working memory of distributed team members and prevents much of the scaffolding often used to reduce memory load in co-located teams (Fiore, Salas, Cuevas, & Bowers, 2003). This research led to the term team opacity, which refers to a special form of workload based on a lack of cues or lack of shared situational awareness.

An accumulating body of research has been undertaken to investigate a host of other distributed teamwork issues (see Bell & Kozlowski, 2002; Zaccaro, Ardison, & Orvis, 2004). For example, Aubert and Kesley (2003) found that trust was less prevalent in distributed teams. Fortunately, distributed teams with low levels of trust were still able to provide high quality outputs. In addition, research has recently been undertaken to address a number of issues related to virtual team leadership (Cascio & Shurygailo, 2002; Zaccaro & Bader, 2003), communication (Jarvenpaa & Leidner, 1999), collaboration (Cohen & Mankin, 1999), conflict (Montoya-Wise, Massey & Song, 2001), and technology (Driskell et al., 2003; Kendall, 1999). Further illuminating the factors that comprise and impinge upon distributed team performance and effectiveness is an agenda that warrants closer consideration in forthcoming research initiatives.

Team Member Individual Differences

In order to achieve and sustain team effectiveness, all organizations must be prepared to identify, import, and invest in human capital. Moreover, there is a growing imperative to

proactively manage human resources at multiple levels (e.g., staffing, selection, training). In a team context the process of securing and developing human capital often involves the assessment of team member characteristics or individual differences. Assessment is conducted in order to quantify the characteristics of individuals within a given population. In team member personnel selection systems the emphasis is on assessing interindividual differences, or those characteristics which differentiate one team member from the next. By contrast, in team member training interventions the concern shifts to the assessment and development of intraindividual differences, or those characteristics which can change within a team member over time.

Individual differences serve as both direct and indirect determinants of teamwork and team performance. In turn, the outcomes of teamwork and team performance are the basis for stakeholder judgments about team effectiveness. Direct determinants of team performance include job relevant knowledge, skill, and volitional choice behavior (Campbell & Kuncell, 2001). Direct determinants are a function of indirect determinants such as cognitive and psychomotor abilities, personality characteristics, perceptions, cognitive styles, cultural assumptions, and experience. In sum, the impact of indirect determinants on team performance is mediated and or moderated by the effort group members exert on a task, the knowledge and skills members can apply to a task, and the task performance strategies used to accomplish a task (Driskell, Hogan, & Salas, 1987).

There is a growing wave of theory building and empirical research concerned with team member individual differences that serve to facilitate team process, performance, and effectiveness (see Barrick, Stewart, Neubert, & Mount, 1998; Barry & Stewart, 1997; LePine, 2003; Neuman & Wright, 1999). The preponderance of research to date has emphasized team member efficacy beliefs, abilities, personality characteristics, and cultural assumptions. By

comparison, however, fewer initiatives have investigated the role of team member cognitive styles, attributions, mental models, and critical thinking skills. This state of affairs is unacceptable if organizations intend to truly leverage the synergies which are possible when a team of experts interdependently interacts as an expert team.

Some lessons learned

Lesson Learned 25: Teams evolve and mature over time (Morgan et al., 1993; Tuckman, 1965).

Lesson Learned 26: As noted earlier, all teams are not created equal; different types of teams have varying functions, structures, processes, and work cycles (Devine, 2002; Sundstrom et al., 2000).

Lesson Learned 27: The complexities of routine and adaptive team performance and effectiveness can be meaningfully captured by integrative theories (Campion et al., 1993; Dickinson & McIntyre, 1997; Hackman, 1987; Pulakos, Arad, Donovan, & Plamondon, 2000).

Lesson Learned 28: A meaningful taxonomy can be developed for describing team performance and functions, emphasizing requirements for coordinated actions of team members to produce synchronized effort (Fleishman & Zaccaro, 1992; Nieva et al., 1978; Shiflett et al., 1985).

Lesson Learned 29: A reliable and valid measurement system for observers to evaluate theperformance of teams on these performance functions can be developed (Brannick et al., 1997; Cooper et al., 1984; Fleishman & Zaccaro, 1992).

Lesson Learned 30: Organizations can navigate complex operations in dynamic environments by utilizing self-managing work teams (Pearce & Ravlin, 1987)

Lesson Learned 31: Teamwork skills are different than taskwork skills (McIntyre & Salas, 1995).

Lesson Learned 32: Multilevel theories of team decision making offer beneficial insight into the process comprising teamwork (Hollenbeck et al., 1995; Ilgen, 1991).

Lesson Learned 33: CRM is an effective (when designed appropriately) and popular team training intervention in aviation, healthcare, and other industries (Helmreich, Weiner, & Kanki, 1993; Salas, et al, 2001).

Lesson Learned 34: Cross-training, team self-correction, team coordination training, scenario-based training, team leadership training, and meta-cognition training are viable instructional strategies for training teams.

Lesson Learned 35: Team training works, helps to promote teamwork, and is vital for effective team performance.

Lesson Learned 36: Teams require measurement that is dynamic, process-oriented, and multilevel (Klein & Kozlowski, 2000). Individual measurements of team members are not enough to capture team level processes and emergent states.

Lesson Learned 37: Shared mental models provide a theoretically grounded platform for understanding team performance in complex environments (Cannon-Bowers et al., 1993; Rentsch & Hall, 1994).

Lesson Learned 38: Distribution in space-time changes aspects of team performance and effectiveness (Bell & Kozlowski, 2002; Fiore et al., 2003; Zaccaro et al., 2004).

Lesson Learned 39: A set of team-based competencies can be outlined and identified (Cannon-Bowers et al., 1995; Stevens & Campion, 1999).

Lesson Learned 40: Team member interindividual and intraindividual differences are important to team effectiveness and can be targeted via selection and training interventions **(**Barrick et al., 1998; LePine, 2003).

With The Benefit of Hindsight…

From this historical account it is clear that teams are integral to the success of individuals, organizations, industries, and whole societies operating in an increasingly complex environment. To date, research has yielded a great deal of information about the utility of teams (see Gibson & Kirkman, 1999). However, much theory building, research, and practice remains to be done. It seems that after compiling all the lessons we have learned over the last century, one message is clear: industries, governments, and organizations count on teams. Even when research is not focused on teams, factories, the military, the medical community, and the aviation industry rely upon teams for operational excellence. Teams have proven to be beneficial and will continue to proliferate as the complexity of our work environments increases. Teams always have and always will exist.

Lesson Learned 41: Teams have stood the test of time….

Lesson Learned 42: Teams are here to stay…

References

Allport, F. H. (1924). *Social Psychology*. Boston: Houghton Mifflin.

Ancona, D. & Chong, C. L. (1999). Cycles and synchrony: The temporal role of context in team behavior. In R. Wageman (Ed.), *Research on Managing Groups and Teams: Groups in Context*. Stamford, Connecticut: JAI Press.

Argote, L. & McGrath, J. D. (1993). Group processes in organisations: Continuity and change. In C.L. Cooper & I.T. Robertson (Eds.), *International Review of Industrial and Organisational Psychology*, (vol. 8). New York: John Wiley & Sons.

Argyris, C. (1970) *Intervention theory and method: A behavioral science view*, Reading, Mass.: Addison Wesley.

Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). *Small groups as complex systems: Formation, coordination, development, and adaptation.* Newbury Park, CA: Sage.

Aubert, B. A. & Kelsey, B. L. (2003). Further understanding of trust and performance in virtual teams. *Small Group Research, 34(5),* 575-618.

Austin, J. R. (2003). Transactive memory systems in organizational groups: The effects of content, consensus, specialization, and accuracy on group performance. *Journal of Applied Psychology, 5*, 866-878.

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology, 3*, 377-391.

Barry, B. & Stewart, G. L. (1997). Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied Psychology, 1,* 62-78.

Bavelas, A. (1960). Leadership: Man and function. *Administrative Science Quarterly, 4*, 491-498.

Bell, B. S. & Kozlowski, S. J. (2002). A typology of virtual teams: Implications for effective leadership. *Group & Organization Management, 27,* 14-49.

Bernstein, P. (1988). The learning curve at Volvo. *Columbia Journal of World Business, 23,* 87-95.

Bettenhausen, K. L. (1991). Five years of group research: What we have learned and what needs to be addressed. *Journal of Management, 17*, 345-381.

Beyerlein, M. & Porter, G. (2000). Historic roots of team theory and practice. In M. Beyerlein (Ed.), *Work teams: Past, present and future*. Dordrecht, The Netherlands: Klewer.

Beyerlein, M., Johnson, D., & Beyerlein, S. (2003), *Team based organizing: Advances in interdisciplinary studies of work teams* (Vol 9). Oxford, UK: Elsevier Science.

Bion, W. R. (1948). Experiences in groups. *Human Relations, vols. I-IV,* 1948-1951.

Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (2000). Analyzing knowledge requirements in team tasks. In J. M. Schraagen and S. F. Chipman (Eds.), *Cognitive task analysis* (pp. 431-447). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Boring, E. (1950). *A history of experimental psychology (2nd ed.).* New York: Appleton Century Crofts.

Bowers, C. A., Braun, C. C., & Morgan, B. B., Jr. (1997). Team workload: Its meaning and measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance and measurement: Theory, methods, and applications,* 85-108. Mahwah, NJ: Lawrence Erlbaum Associates.

Brannick, M. T., Salas, E., & Prince, C. (Eds.). (1997). *Team performance assessment and measurement: Theory, methods, and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Briggs, G. E. & Johnston, W. A. (1967). *Team training* (NAVRADEVCEN-1327-4, AD-660019). Orlando, FL: Naval Training Device Center.

Campbell, J. P. (1990). The role of theory in Industrial and Organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (pp.39-74). Palo Alto, California: Consulting Psychologists Press.

Campbell, J. P. & Kuncel, N. R. (2001). Individual and team training. In N. Anderson, D. S. Ones, Sinangil, H. K., and C. Viswesvaran (Eds.), *Handbook of Industrial, Work and Organizational Psychology: Personnel Psychology, Vol. 1* (pp. 278 – 312). London: Sage Publications.

Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology, 46,* 823–850.

Campion, M. A., Papper, E. M., & Medsker, G. J. (1996). Relations between work team characteristics and effectiveness: A replication and extension. *Personnel Psychology, 49,* 429-452.

Cannon-Bowers J. A. & Salas E. (1997). A framework for developing team performance measures in training. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, research, and applications* (pp.45-62). Mahwah, NJ: Lawrence Erlbaum Associates.

Cannon-Bowers, J.A. & Salas, E. (Eds.), (1998). *Decision making under stress: Implications for training and simulation*. Washington, DC: American Psychological Association.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision-making. In N. J. Castellan Jr. (Ed.), *Individual and group decision making* (pp. 221-246). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cascio, W. F. & Shurygailo, S. (2002). E-Leadership and virtual teams. *Organizational Dynamics, 31,* 362-376.

Chen, G. & Ployhart, R. E. (2004). *Time for a change: Towards a theory of performance change.* Presented at the 19th Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Cohen, S. G. & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management, 23*, 239-290.

Cohen, S. G. & Mankin, D. (1999). Collaboration in the virtual organization. In C. L. Cooper and D. M. Rousseau (Eds.), *Trends in organizational behavior, Vol. 6: The virtual organization* (pp. 105-120). Chichester, England: John Wiley & Sons Ltd.

Cohen, S., Ledford, G., & Spreitzer, G. (1996). A predictive model of self-managing work team effectiveness. *Human Relations, 49*, 643-676.

Cooper, M., Shiflett, S., Korotkin, A. L., & Fleishman, E. A. (1984). *Command and control teams: Techniques for assessing team performance* (AFHRL-TP-84-3). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Cranton, P. (1996). Types of group learning. In S. Imel (Ed.), *Learning in groups: Exploring fundamental principles, new uses, and emerging opportunities. New directions for adult and continuing education* (pp. 25-32). San Francisco: Jossey-Bass.

Davis, J. H. (1992). Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples, 1950-1990. *Organizational Behavior and Human Decision Processes, 52*, 3-38.

Day, D. V. & Lance, C. E. (2004). Modeling leader growth and development. In D. V. Day, S. J. Zaccaro, & S. M. Halpin (Eds.) *Leadership development for transforming organizations* (pp. 41-69). Mahwah, NJ: Erlbaum.

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004, December). Multiple goal feedback effects on the regulation of individual and team performance in training. *Journal of Applied Psychology, 89(6),* 1035-1056.

Devine, D. J. (2002). A review and integration of classification systems relevant to teams in organizations. *Group Dynamics: Theory, Research, and Practice, 6*, 291-310.

Devine, D. J., Clayton, L. D., Philips, J. L., Dunford, B. B., & Melner, S. B. (1999). Teams in organizations: Prevalence, characteristics and effectiveness. *Small Group Research, 30*, 678-711.

Dickinson, T. L. & McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. In M. T. Brannick, E. Salas, and C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 19-43). Mahwah, NJ: Erlbaum.

Driskell, J. E., Hogan, R., & Salas, E. (1987). *Personality and group performance. Group processes and intergroup relations: Review of personality and social psychology* (pp. 91-112). Beverly Hills, CA: Sage Publications.

Driskell, J. E., Radtke, P. H., & Salas, E. (2003). Virtual teams: Effects of technological mediation on team performance. *Group Dynamics, 7,* 297-323.

Drucker, P. (1992). *The age of discontinuity: Guidelines to our changing society*. Piscataway, N.J.: Transaction Publishers.

Durkheim, E. (1895). *Rules of sociological method.* New York: Free Press.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44, 350-383.

Ensley, M. D. & Pearce, C. L. (2001). Shared cognition in top management teams: Implications for new venture performance. *Journal of Organizational Behavior, 22,* 145-160.

Entin, E. E. & Serfaty, D. (1999). Adaptive team coordination. *Human Factors, 41,* 312-325.

Federal Aviation Administration (FAA) (1998). *Crew Resource Management Training* (Advisory Circle No: 120-51C). Available at http://www.crm-devel.org/resources /ac/ac120_51c.html.

Fiore, S. M., Salas, E., Cuevas, H. M., & Bowers, C. A. (2003). Distributed coordination space: toward a theory of distributed team process and performance. *Theoretical Issues in Ergonomics Science, 4,* 340-364.

Fleishman, E.A. (1953). The description of supervisory behavior. *Personnel Psychology, 37,* 1-6

Fleishman, E.A. (1975). Toward a taxonomy of human performance. *American Psychologist, 30,* 1127-1149.

Fleishman, E. A. & Quaintance, M. K. (1984). *Taxonomies of Human Performance*. Orlando: Academic Press.

Fleishman, E. A. & Zaccaro, S. J. (1992). Toward a taxonomy of team performance functions. In R.W. Swezey, & E. Salas (Eds.), *Teams: Their training & performance*. Norwood, NJ: Ablex.

Flin, R. (1995). Crew resource management for teams in the offshore oil industry. *Journal of European Industrial Training, 19,* 23-27.

Follett, M. P. (1926). The giving of orders. *COT*, 156-162.

Ford, J.K. (1997). Advances in training research and practice: An historical perspective. In J. K. Ford & Associates (Eds.). *Improving training effectiveness in work organizations,* (pp. 1 - 18), LEA.

Frost, C. F. (1974). *The Scanlon plan for organization development: Identity, participation, and equity*. Lansing, MI: Michigan State University Press.

Gaba D. M., Howard S. K., & Fish K. J. (1994). *Crisis Management in Anesthesiology.* New York, Churchill Livingstone.

Gersick, C. J. (1989, June). Marking time: Predictable transitions in task groups. *Academy of Management Journal, 32(2),* 274-309.

Gersick, C. G. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal, 31,* 9-41.

Gibson, C.B., & Kirkman, B.L. (1999). Our past, present, and future in teams: The role of human resources professionals in managing team performance. In A. I. Kraut & A. K. Korman (Eds.), *Evolving practices in human resources management: Responses to a changing world of work* (pp. 90-117). San Francisco: Jossey-Bass.

Gladstein, D. L. (1984). Groups in context: A model of task group effectiveness. *Administrative Science Quarterly, 29,* 499-517.

Glassop, L. I. (2002). The organizational benefits of teams. *Human Relations, 55,* 225-249.

Glickman, A. S., Zimmer, S., Montero, R. C., Guerette, P. J., Campbell, W. J., Morgan, B. B., Jr., & Salas, E. (1987). *The evolution of teamwork skills: An empirical assessment with*

implications for training (NTSC Tech. Rep. No. 87-016). Orlando, FL: Naval Training Systems Center.

Griffith, T. L. & Neale, M.A. (2001). Information processing in traditional, hybrid, and virtual teams: From nascent knowledge to transactive memory. In B. Staw and R. Sutton (Eds.), *Research in organizational behavior, Vol. 23*. Stamford, CT: JAI Press.

Guzzo, R. A. & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology, 47*, 307-338.

Guzzo, R. A. & Shea, G. P. (1992). Group performance and intergroup relations in organizations. In M. D. Dunnette and L. M. Hough (Eds.), *Handbook Of Industrial And Organizational Psychology* (2nd edition). Palo Alto, CA: Consult. Psych. Press.

Gyllenhammer, P. (1973). Volvo's solution to the blue-collar blues. *Business and Society/Innovation, 7*, 50-54.

Hackman, J. R. (1987). The design of work teams. In J. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315-342). New York: Prentice Hall.

Hackman, J. R. (1990). *Groups that work (and those that don't).* San Francisco: Jossey-Bass.

Hackman, J. R., & Lawler, E. E. (1971). Employees' reactions to job characteristics. *Journal of Applied Psychology, 55*, 259-285.

Hackman, J. R. & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology*, 60, 159-70.

Harrison, D. A., Mohammed, S., McGrath, J. E., Florey, A. T., & Vanderstoep, S. W. (2003). Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality. *Personnel Psychology, 56*, 633-669.

Highhouse, S. (2002a). Assessing the candidate as a whole: An historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology, 55,* 363-396.

Highhouse, S. (2002b). A history of the T-group and its early applications in management development. *Group Dynamics, 6,* 277-290.

Hollenbeck, J. R., Ilgen, D. R., LePine, J. A., Colquitt, J. A., & Hedlund, J. (1998). Extending the multilevel theory of team decision making: Effects of feedback and experience in hierarchical teams. *Academy of Management Journal, 41,* 269-282.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology, 80,* 292-316.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *International Journal of Aviation Psychology, 9,* 19-32.

Helmreich, R. L., Wiener, E. L., & Kanki, B. G. (1993). The future of crew resource management in the cockpit and elsewhere. In E. L. Wiener, B. G. Kanki, et al. (Eds), *Cockpit resource management* (pp. 479-501). San Diego, CA: Academic Press, Inc.

Hemphill, J.K. (1949). *Situational factors in leadership*. Columbus, Ohio: Ohio State University.

Herzberg, F. (1966). *Work and the nature of man.* Cleveland: World Publishing Co.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121,* 43-64.

Ilgen, D. R. (1999). Teams in organizations: Some implications. *American Psychologist, 54,* 129-139.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From Input-Process-Output Models to IMOI models. *Annual Review of Psychology, 56,* 517-543.

Ilgen, D. R., Major, D. A., Hollenbeck, J. R., & Sego, D. J. (1991). *Decision making in teams: Raising an individual model to the team level* (Technical Report No. 91-92). East Lansing: Michigan State University.

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology, 10,* 371-384.

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes.* Boston, MA: Houghton Mifflin.

Jarvenpaa, S. L. & Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization Science, 10,* 791-815.

Katzenback, J. R. & Smith, D. K. (1993). The discipline of teams. *Harvard Business Review, 7,* 111-120.

Kendall, K. E., (Ed) (1999). *Emerging information technologies: Improving decisions, cooperation, and infrastructure.* Camden, NJ: Rutgers University.

Klein, G. A. (1998). *Sources of power: how people make decisions.* Cambridge, MA: MIT Press.

Klein, G. & Pierce, L. (2001). Adaptive teams. (Draft report, purchase Order H438556 (A) for Link Simulation and training Division/Army Prim Contract no. DAAD17-00-A-5002). Fairborn, OH: Klein Associates.

Klein, K. & Kozlowski, S. W. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein and S. W. J.

Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3-90). San Francisco: Jossey-Bass, Inc.

Kleinman, D. L. & Serfaty, D. (1989). *Team Performance assessment in distributed decision making*. Proceedings of the Symposium on Interactive Networked Simulation for Training (pp. 22-27), Orlando, FL.

Klimoski, R. & Mohammad, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20,* 403-437.

Kozlowski, S. W. J. & Bell, B. S. (2002). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Comprehensive handbook of psychology (Vol. 12): Industrial and Organizational Psychology*. New York: Wiley.

Kozlowski, S. W. J., Gully, S. M., Brown, K. G., Salas, E., Smith, E. A., & Nason, E. R. (2001). Effects of training goals and goal orientation traits on multi-dimensional training outcomes and performance adaptability. *Organizational Behavior and Human Decision Processes, 85,* 1-31.

Kozlowski, S. W. J, Gully, S. M., Nason, E. R., & Smith, E. M. (1999). Developing adaptive teams: A theory of compilation and performance across levels and time. In D. R. Ilgen and E. D. Pulakos (Eds.), *The changing nature of work and performance: Implications for staffing, personnel actions, and development* (First ed., pp.240-292). San Francisco, CA: Jossey-Bass.

Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development systems in organizations: State-of-the-art lessons for practice* (pp. 331-375). San Francisco, CA: Jossey-Bass.

LePine, J. A. (2003). Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. *Journal of Applied Psychology, 88,* 27-39.

Levine, J. M. & Moreland, R. L. (1991). Culture and socialization in work groups. In L. B. Resnick and J. M. Levine (Eds.), *Perspectives on socially shared cognition* (pp. 257-279). Washington, DC: American Psychological Association.

Lewin, K. (1939). Field theory and experiment in social psychology: Concepts and methods. *American Journal of Sociology, 44,* 868-896.

Lewin, K., Lippitt, R., & White, R.K. (1939). Patterns of aggressive behavior in experimentally created "social climates." *Journal of Social Psychology, SPSSI Bulletin, 10,* 271-299.

Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology, 88(4),* 587-604.

Management Review, (1972). The plant that runs on individual initiative, *61,* 20-26.

Manz, C. C. & Neck, C. P. (1995). Teamthink: Beyond the group think syndrome in self-managing work teams. *Journal of Managerial Psychology, 10,* 7-15.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team process. *Academy of Management Review, 26,* 356-376.

Marks, M.A., Zaccaro, S. J., & Mathieu, J. E. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *Journal of Applied Psychology*, 6, 971-986.

Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology, 87,* 3-13.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50,* 370-396.

Mathieu, J. E., Heffner, T. S., & Goodwin, G. F. (1996). *Linking cognitive and behavioral components of aircrew coordination.* Naval Air Warfare Center, Orlando, FL.

McCarthy, P. (2002). *Brief outline of the history of I/O Psychology.* Retrieved January 21, 2002, from http://www.mtsu.edu/~pmccarth/io_hist.htm

McDougall, W. (1930). William McDougall. In G. Lindzey (Ed.), *A history of psychology in autobiography, Vol. 7*. Worcester, MA: Clark University Press.

McGrath, J. (1964). *Social Psychology: A brief introduction.* New York, NY: Holt, Rinehart, and Winston.

McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The study of groups: Past, present, and future. *Personality and Social Psychological Review, 4*, 95-105.

McGregor, D. M. (1957). The human side of enterprise. *Management Review, 46*, 88-92.

McIntyre, R. M. & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 9-45). San Francisco: Jossey-Bass.

McKeachie, W. J. & Brim, O. G. (1984). Lessons to be learned from large behavioral research organizations. *American Psychologist, 11*, 1254-1255.

Merritt, A. C. & Helmreich, R. L. (1995). CRM in 1995: Where to from here? In B. J. Howard and A. R. Lowe (Eds.), *Applied aviation psychology: Achievement, change, and challenge. Proceedings of the 3rd Australian Aviation Psychology Symposium* (pp. 111-126). Aldershot, England: Avebury Aviation.

Moede,W. (1927). Die Richtlinien der Leistungs-Psychologie. *Industrielle Psychotechnik, 4*, 193–207.

Mohrman, S., Cohen, S., & Mohrman, A. (1995). *Designing team-based organizations: New forms for knowledge work.* San Francisco: Jossey Bass.

Montoya-Weiss, M. M., Massey, A. P., & Song, M. (2001). Getting it together: Temporal coordination and conflict management in global virtual teams. *Academy of Management Journal, 44*, 1251-1262.

Moreland, R. L. (1999). Transactive memory: Learning who knows what in work groups and organizations. In L. L. Thompson and J. M. Levine (Eds.), *Shared cognition in organizations: The management of knowledge* (pp. 3-31). Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, B., Salas, E., & Glickman, A. S. (1993). An analysis of team evolution and maturation. *Journal of General Psychology, 120,* 277-291.

Mullen, B. & Cooper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin, 115,* 210-227.

Neuman, G. A. & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology, 84*, 376-389.

Nieva, V. F., Fleishman, E. A., & Rieck, A. M. (1978). *Team dimensions: Their identity, their measurement, and their relationships*. Technical report. Washington, DC: ARRO.

Orasanu, J. (1990). *Shared mental models and crew performance*. Paper presented at the 34th Annual Meeting of the Human Factors Society. Orlando, FL.

Orasanu, J. & Salas, E. (1993). Team decision making in complex environments. In G. A. Klein, J. Orasanu, et. al. (Eds.), *Decision making in action: Models and methods* (pp. 327-345). Stamford, CT: Ablex Publishing Corp.

Pearce, J. A. I. & Ravlin, E. C. (1987). The design and activation of self regulating work groups. *Human Relations, 40*(11), 751-760.

Peters, T. & Waterman, J. (1982). *In search of excellence*. New York: Warner Books.

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptation in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 85,* 612-624.

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptation in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 85,* 612-624.

Pulakos, E. D., Schmitt, N., Dorsey, D. W., Arad, S., Hedge, J. W., & Borman, W. C. (2002). Predicting adaptive performance: Further tests of a model of adaptability. *Human Performance, 4,* 299-323.

Rentsch, J. R. & Hall, R. J. (1994). Members of great teams think alike: A model of team effectiveness and schema similarity among team members. In M. M. Beyerlein and D. A. Johnson (Eds.), *Advances in interdisciplinary studies of work teams: Theories of self-managing work teams, Vol. 1* (pp. 223-261). US: Elsevier Science/JAI Press.

Roby, T. B. & Lanzetta, J. T. (1957). Conflicting principles in man-machine system design. *Journal of Applied Psychology, 41,* 170-178.

Roethlisberger, F. J. (1941). The Hawthorne experiments. *Management and morale*. Cambridge, Massachusetts: Harvard University Press.

Rogers, C. R. (1970). *Carl Rogers on encounter groups*. New York: Harper Collins.

Salas, E., & Cannon-Bowers, J. A. (2000). Designing training systems systematically. In E. A. Locke (Ed.), *The Blackwell handbook of principles of organizational behavior* (pp. 43-59). Malden, MA: Blackwell Publisher Ltd.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52,* 471-499.

Salas, E. & Fiore, S. M. (Eds.) (2004). *Team cognition: Understanding the factors that drive process and performance.* Washington, DC: APA.

Salas, E., Bowers, C. A., & Cannon-Bowers, J. A. (1995). Military team research: Ten years of progress. *Military Psychology, 7,* 55-75.

Salas, E., Bowers, C. A., & Edens, E. (2001). An overview of resource management in organizations: Why now? In E. Salas, C. A. Bowers, & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 1-5). Mahwah, NJ: Lawrence Erlbaum Associates.

Salas, E., Stagl, K. C. & Burke, C. S. (2004). 25 years of team effectiveness in organizations: Research themes and emerging needs. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology*. New York: John Wiley & Sons.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey and E. Salas, (Eds.), *Teams: Their training and performance* (pp. 3-29). Stamford, CT: Ablex Publishing Corp.

Salas, E., Rozell, D., Mullen, B., & Driskell, J. E. (1999). The effect of team building on performance: An integration. *Small Group Research, 30,* 309-329.

Salas, E., Stagl, K. C., Burke, C. S., & Goodwin, G. F. (in press). Fostering team effectiveness in organizations: Toward an integrative theoretical framework of team performance. In J. W. Shuart, W. Spaulding, & J. Poland, (Eds.), *Modeling complex systems: Motivation, cognition and social processes, Nebraska Symposium on Motivation, 51*, Lincoln, NE: University of Nebraska Press.

Salas, E., Wilson, K. A., Burke, C. S., & Bowers, C. A. (2002). Myths about crew resource management training. *Ergonomics in Design,* 20-24.

Shafritz, J. M. & Ott, J. S. (Eds.) (2001). *Classics of organization theory, 5th ed.* San Diego: Harcourt.

Sheard, A. & Kakabadse, A. (2002). From loose groups to effective teams: The nine key factors of the team development landscape. *Journal of Management Development, 2,* 131-151.

Shiflett, S., Eisner, E. J., Price, S. J., & Schemmer, F. M. (1985). The definition and measurement of military small unit team functions. Research report. Bethesda, MD: ARI.

Steiner, I. (1972). *Group process and productivity.* New York: Academic Press.

Stevens, M. J. & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management, 25(2),* 207-228.

Stewart, G., Manz, C., & Sims, H. (1999). *Team work and group dynamics*. New York: Wiley and Sons.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors, 41*, 61-71.

Sundstrom, E. D., De Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist, 45,* 120-133.

Sundstrom, E., McIntyre, M., Halfhill, T., & Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics, 4*, 44-67.

Swezey, R. W. & Salas, E. (1992). Guidelines for use in team-training development. In R. W. Swezey and E. Salas (Eds.), *Teams: Their training and performance* (pp. 219-245). Westport, CT: Ablex Publishing.

Tang, T. L. & Butler, E. A. (1997). Attributions of quality circles' problem-solving failure: Differences among management, supporting staff, and quality circle members. *Public Personnel Management, 26,* 203-225.

Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology, 43,* 399-441.

Tannenbaum, S. I., Beard, R. L., & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. In K. Kelley (Ed.), *Issue, theory, and research in Industrial/Organizational Psychology* (pp. 117-153). Amsterdam: Elsevier.

Taylor, F. W. (1911). *The principles of scientific management.* New York: Norton.

Tedford, K. & Baker, D. (2000). Kurt Lewin: Contributions to organizational development. In M. Beyerlein (Ed.) *Work teams: Past, present and future*. Dordrecht, The Netherlands: Klewer.

Tesluk, P., Mathieu, J. E., Zacarro, S. J., & Marks, M. (1997). Task and aggregation issues in the analysis and assessment of team performance. In M. T. Brannick, E. Salas, & C. Prince

(Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 197-224). Mahwah, NJ: Erlbaum.

Townsend, A. M., DeMarie, S. M., & Hendrickson, A. R. (1996, September). Are you ready for virtual teams? *HR Magazine, 41*, 122-128.

Trist, E. L., Emory, F., Murray, H., & Trist, B. (Eds.) (1997). The social engagement of social science: A Tavistock anthology: The socio-ecological perspective. Philadelphia: University of Pennsylvania Press.

Tuckman, B. W. (1965). Personality, structure, group composition, and group functioning. *Sociometry, 27,* 469-487.

Waller, M. (1999). The timing of adaptive group responses to nonroutine events. *Academy of Management Journal, 42*, 127-137.

Watson, J.B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20,* 158-177.

Weber, M. (1972). Bureaucracy. In H. Gerth & C. W. Mills (Eds.) *Max Webber: Essays in sociology*. Oxford, UK: Oxford University Press.

Wegner, D. (1987). Transactive memory: a contemporary analysis of the group mind. In B. Mullen & G. Goethals (Eds.), *Theories of group behavior* (pp. 185-208). New York: Springer-Verlag.

Weingart, L. R. (1997). How did they do that? The ways and means of studying group prcess. *Research in Organizational Behavior, 19,* 189-239.

West, M. A., Tjosvold, D., & Smith, K. G. (2003). *International handbook of organizational teamwork and cooperative working*. West Sussex, England: John Wiley & Sons.

Wheelan, S. A. (1994). *Group processes: A developmental perspective.* Needham Heights, MA: Allyn & Bacon,

Wellins, R. S., Byham, W. C., & Wilson, J. M. (1991). *Empowered teams: creating self-directed work groups that improve quality, productivity, and participation.* San Francisco: Jossey-Bass Publishers.

Whitsett, D. A., & Yorks, L. (1983). Looking back at Topeka: General Foods and the quality-of-work-life experiment. *California Management Review, 25*, 93-109.

Zaccaro, S. J. & Bader, P. (2003). E-Leadership and the challenges of leading E-teams: Minimizing the bad and maximizing the good. *Organizational Dynamics, 31,* 377-387.

Zaccaro, S. J., Ardison, S. D., & Orvis, K. L. (2004). Leading virtual teams. In D. Day, S. J. Zaccaro, & S. M. Halpin (Eds.), *Leader development for transforming organizations*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Acknowledgements

The authors would like to acknowledge Drs. Richard Guzzo, Scott Highhouse, Laura Koppes, Paul Thayer, and 3 anonymous historian reviewers for their editorial feedback and insightful comments. We are particularly thankful to Ed Fleishman for providing numerous historical sources and information about teams.

Figure 1. History of Teams Timeline

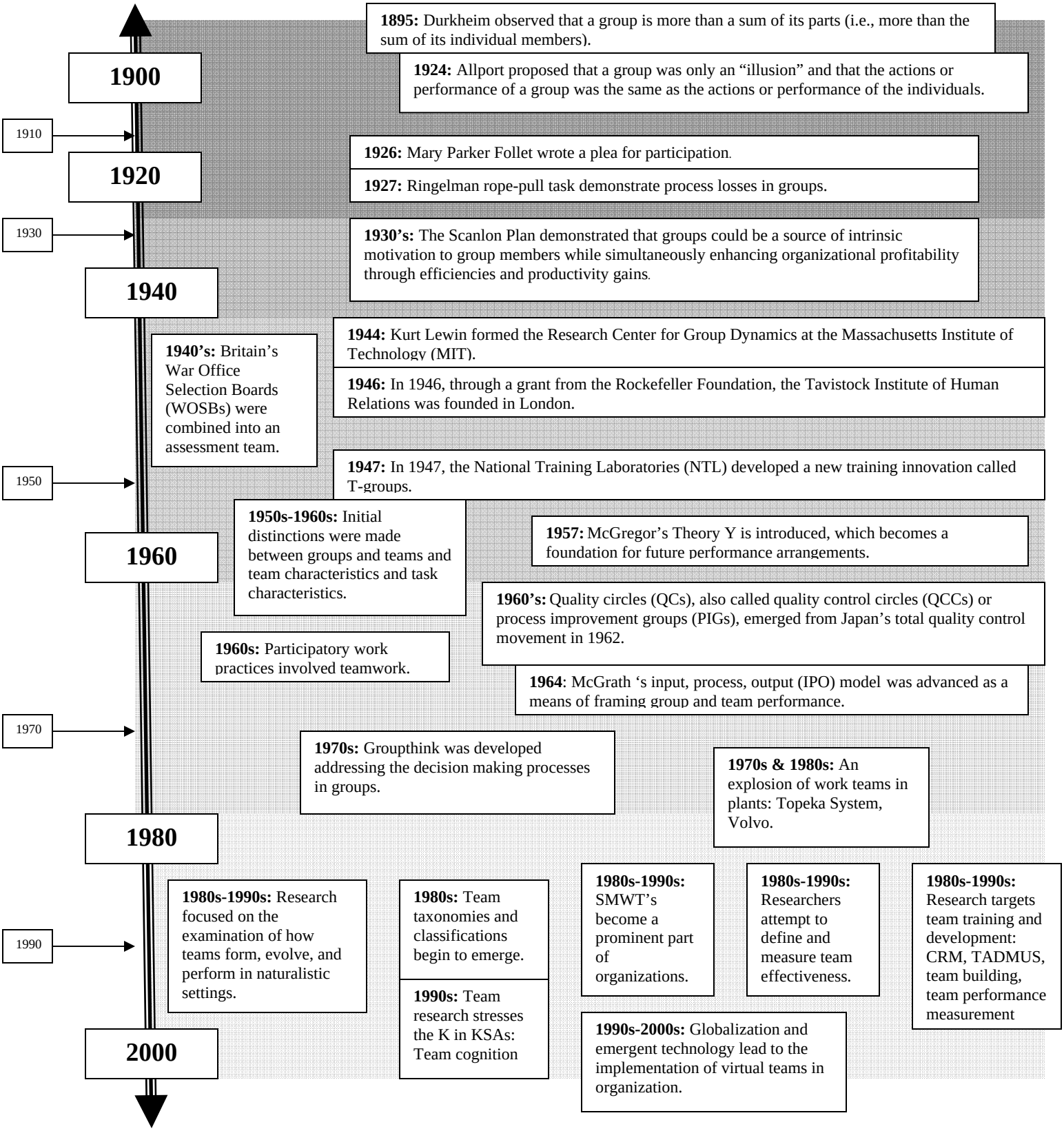


Table 1

Commonly Examined Characteristics of Groups (Hemphill, 1949)

Group Dimensions	**Dimension Description**
Group Size	Number of members in the group
Viscidity	Group characteristics that differentiates a group from a collection of individuals
Homogeneity	Diversity in membership (e.g., age, gender)
Flexibility	Degree group adheres to a standard set of behaviors
Permeability	Restriction on those who are considered group members or able to join the group
Polarization	Orientation towards a single, specific goal
Stability	Pace change occurs within the group
Intimacy	Degree of association among members
Autonomy	Independence from other groups
Control	Amount of behavior regulation by other team members
Position	Role or status within the group
Potency	Individual member's need satisfaction by group membership
Hedonic Tone	Feelings of satisfaction, pleasantness, or agreeableness associated with group membership
Participation	Amount of time/effort spent on group tasks
Dependence	Relationship between the team and its leader in ensuring group needs are met

Table 2

A Taxonomy of Team Functions*

I. Orientation Functions
- a. Information Exchange Regarding Member Resources and Constraints
- b. Information Exchange Regarding Team Task and Goals/Mission
- c. Information Exchange Regarding Environmental Characteristics and Constraints
- d. Priority Assignment Among Tasks

II. Resource Distribution Functions
- a. Matching Member Resources to Task Requirements
- b. Load Balancing

III. Timing Functions (Activity Pacing)
- a. General Activity Pacing
- b. Individually Oriented Activity Pacing

IV. Response Coordination Functions
- a. Response Sequencing
- b. Time and Position Coordination of Responses

V. Motivational Functions
- a. Development of Team Performance Norms
- b. Generating Acceptance of Team Performance Norms
- c. Establishing Team-Level Performance-Rewards Linkages
- d. Reinforcement of Task Orientation
- e. Balancing Team Orientation with Individual Competition
- f. Resolution of Performance-Relevant Conflicts

VI. Systems Monitoring Functions
- a. General Activity Monitoring
- b. Individual Activity Monitoring
- c. Adjustment of Team and Member Activities in Response to Errors and Omissions

VII. Procedure Maintenance
- a. Monitoring of General Procedural-Based Activities
- b. Monitoring of Individual Procedural-Based Activities
- c. Adjustments of Nonstandard Activities

*From Fleishman, E.A. & Zacarro, S.J. (1992)

Assessment of Team Learning

Kevin C. Stagl, Eduardo Salas

Department of Psychology

&

Institute for Simulation & Training

University of Central Florida

&

David V. Day

Department of Psychology

Pennsylvania State University

DO NOT CITE OR CIRCULATE WITHOUT PERMISSION OF THE AUTHORS

Address correspondence to: Eduardo Salas
Institute for Simulation & Training
University of Central Florida
3100 Technology Parkway
Orlando, FL 32826
(407) 882-1325 (w); 658-5059 (fax)
esalas@ist.ucf.edu

Stagl, K. C., Salas, E., & Day, D. V. (in progress). Assessment of team learning. In V. Sessa & M. London (Eds.), Group learning (pp.). Mahwah, NJ: Lawrence Erlbaum and Associates.

Assessment of Team Learning

During the first half of gestation, hundreds of millions of neurons form an intricate web within a 'dynamic learning machine', says Dr. Carla Shatz, Chair of the Neurobiology Department at Harvard Medical School (www.dana.org). The machine, of course, is a baby, and as its neural network mushrooms its nervous system is busy at work checking and rechecking established connections. At this early stage, genetics largely dictate the complex connections being made; but soon after the balance of influence shifts to nurture and via learning a toddler continuously undergoes changes in the synapses or linkages between neurons. Over time, learning leads to biochemical and possibly genetic changes that, in turn, give rise to memory.

While amazing in its own right, the brain of a child is qualitatively different than that of an adult team member. For example, the frontal lobes, which are critical for organizational tasks, do not reach maturation until year 20. Also, in parts of the adult brain like the hippocampus, generic stem cells can miraculously morph into new neurons via the process of neurogenesis. The news is not all positive, however, as even the most health conscious reader has already lost about half of the synaptic pathways that they commanded at age 2. Moreover, the plasticity of these connections erodes over time, making it difficult to forge new pathways in adulthood.

The neurosciences have undoubtedly advanced our understanding about the genetic, cellular, and molecular mechanisms that underlie growth and decline in 'dynamic learning machines'. Unfortunately, this body of research provides less practical insight at other levels of analysis such as those pertinent to designing a suite of assessment tools that can be used to measure, track, and manage the learning process and its outcomes in work settings. Yet, assessing team learning and its effects in the workplace is important because the use of teams as a preferred performance arrangement is often predicated upon the assumption that collectives can

develop and draw upon a broader array of competencies and response processes to address the complex problems which typify modern operations (Ilgen, Hollenbeck, Johnson, & Jundt, 2005).

When it comes to scaling learning and its effects in the workplace, the theories and findings of the behavioral sciences are paramount. According to this body of science, teams are often used as a preferred performance arrangement to structure work because their members can lean on and learn from one another before, during, and after task performance episodes. Thus, team learning is both a precursor to, and an outcome of, routine team performance and effectiveness judgments. In fact, research findings suggest learning is vital to the performance of a wide range of teams including: research and development teams in the energy industry (Van Der Vegt & Bunderson, 2005), enterprise resource planning system implementation teams (Yeh & Chou, 2005), command and control project teams (Ellis, Hollenbeck, Ilgen, Porter, West, & Moon, 2003), and uninhabited air vehicle control teams (Cooke, Kiekel, & Helm, 2001).

Team learning is also an input into and an outcome of adaptive team performance. For example, a recently advanced theory of team adaptation frames team learning as one core component of a recursive cycle of adaptive team performance that ultimately culminates in team adaptation (Burke, Stagl, Salas, Pierce, & Kendall, in press). Qualitative research findings support this line of thinking, suggesting team learning was an important driver of the adaptation of routines that ultimately resulted in the adoption of an innovative technology by cardiac surgery teams (Edmondson, Bohmer, & Pisano, 2001). Moreover, recent empirical evidence suggests learning is an important driver of adaptive performance at both the individual member and team levels of analysis (Chen, Thomas, & Wallace, 2005).

This chapter leverages some of the theories, measurement techniques, and lessons learned in the behavioral sciences to advance a set of assessment tools that can be used to

measure learning and its outcomes in teams embedded in organizations. We begin by distinguishing the process of learning from its outcomes, discussing its emergent nature, and by bracketing this phenomenon in a hierarchical nomological network. Next, a sample of learning outcomes are presented which are often operationalized in terms of learning measures in the workplace. Finally, a set of assessment tools is advanced which can be utilized by practitioners to measure growth and decline in 'dynamic learning teams'.

The Process of Team Learning

Team learning is a process whereby knowledge, skills, and affective states are acquired via dyadic interactions and shared experience (Kozlowski & Ilgen, in press). The process of team learning is multifaceted, manifesting in a myriad of specific behaviors such as: information sharing, joint reflection, asking questions, seeking feedback, experimentation, the discussion of errors and differences, testing assumptions, forming new routines, and adjusting strategies (Edmondson, 1999; Edmondson et al., 2001). When team members learn, they experience growth in the form of differentiation from and integration with the systems in which they are embedded. Teams "…at higher levels of development are able to use a greater number of knowledge principles to construct their experiences (differentiation), make more interconnections among these principles (integration). This results in a broader perspective…" (Day & Lance, 2004, p.43). Thus, learning activities culminate in an increased capacity to engage in the transition, action, and interpersonal processes that comprise team performance and lead to valued team outcomes (see Marks, Mathieu, & Zaccaro, 2001).

Team learning is a team level phenomenon; distinct in nature from both individual (Argyris & Schön, 1978; Halpern, 2004) and organizational learning (Levitt & March, 1988). As a team level property rooted in individual learning, team learning can emerge in a number of

qualitatively different ways (Kozlowski & Klein, 2000). For example, team learning can manifest at the team level via a convergent emergence process, which is a type of isomorphic composition. Composition processes describe the coalescence of identical lower level properties that yield a similar higher level property. Learning can be conceptualized at the team level via a convergent emergence model if, for example, the three members of a team all engaged in an identical set of learning behaviors; such as would be the case if each member shared equally relevant ideas about implementing a new standard operating procedure.

In most work settings, however, team learning is not a process of synchronized, identical actions; but rather emerges from the unique configuration or pattern of social interactions and dyadic exchanges that characterize efforts to navigate complex challenges. A configural compilation emergence model is based on the assumptions of discontinuity, whereby functionally equivalent, but not identical, phenomenon unfold over time to compile across levels (Kozlowski & Klein, 2000). For example, one member of a sales support team may be actively experimenting with various strategies of influence with a senior representative of their team's client, while, concurrently, three other team members share information with a second client rep about pricing points. This kind of compatible, yet distinct, set of actions can culminate in common ground about what tactics and services are desirable for use with a given client.

The above noted examples suggest the process of learning can be conceptualized at the team level in a variety of different ways. Similarly, the outcomes of team learning can manifest at the team level via unique emergence processes. For example, knowledge often becomes embedded in a team as a result of learning activities but the nature of this knowledge can change depending upon the specific learning behaviors enacted, the communication of information within a team, and/or contextual constraints. In the first example noted above, all team members

were involved in a joint discussion of a major error that resulted in the establishment of a new standard operating procedure. Because the team possessed shared knowledge of the new policy at the conclusion of the meeting, as a collective they held a shared mental model (see Klimoski & Mohammed, 1994) of how the new policy would shape future interactions in regards to this issue. Conversely, the second example illustrates how a team might develop a compatible mental model, or unique knowledge held by individual team members which forms a congruent whole (Kozlowski, Gully, Salas, & Cannon-Bowers, 1996; Kozlowski & Klein, 2000).

The most important point raised in the above examples is that the elemental contributions of individual level activities and outcomes have important implications for how team level phenomena manifest and are ultimately measured in the workplace. Mapping the myriad of paths via which team learning and its outcomes can emerge is beyond the scope of this chapter; but it should be noted that the manner in which a particular emergence process is conceptualized is directly relevant to subsequently selecting elicitation methods, combination rules, and metrics to measure learning and its outcomes at the team level. This speaks to the importance of aligning constructs, models of emergence, metrics, and ultimately the research designs and analyses that are employed to generate construct validity evidence in support of the interpretations and actions taken on the basis of psychological assessments (Kozlowski & Klein, 2000).

The remainder of this chapter is divided into two sections. The next section introduces a set of outcomes that result from team learning. A discussion of learning outcomes may seem counterintuitive, given that team learning was framed as a behavioral process previously in this chapter. The emphasis on team learning outcomes rather than the actual activities underpinning learning is purposeful, however, as most attempts to measure learning in the workplace infer team learning from changes in team outcomes rather than via the direct measurement of the team

learning construct (see Kozlowski & Ilgen, in press). This chapter seeks to incrementally inform current organizational practice, while noting that additional theory building and testing is warranted to continue to develop reliable process measures of team learning.

The Outcomes of Team Learning

The unfolding behavioral process of team learning is rarely assessed directly (see Edmondson, 1999; Ellis et al., 2003 for exceptions). Rather, the outcomes of growth are often measured as proxies for team learning (Kozlowski & Ilgen, in press). The specific outcomes of team learning discussed in this chapter were selected for review because they are consistent with theories and frameworks of team performance and development (see Burke et al., in press; Campion, Medsker, & Higgs, 1993; Dickinson & McIntyre, 1997; Hackman, 1987; Marks et al., 2001; Morgan, Salas, & Glickman, 1994; Salas, Stagl, Burke, & Goodwin, in press; Tannenbaum, Beard, & Salas, 1992). For example, Burke and colleagues' theory of team adaptation models the shared cognitive states that emerge from the process of team learning such as shared mental models, team situation awareness, and psychological safety.

Many of the team learning outcomes discussed in this chapter are also consistent with taxonomic efforts undertaken to classify the outcomes of developmental initiatives (see Kraiger, 2002; Kraiger, Ford, & Salas, 1993). Kraiger and colleagues' suggest learning outcomes can be classified into three broad taxons (affective, behavioral, and cognitive), the *ABCs of Learning*. These taxonomies are grounded in learning research that has been conducted in the field of education since the 1950s (see Bloom, 1956), but also draw heavily upon recent instructional design principles (see Jonassen & Tessmer, 1996). Together, the theories of team performance noted above, and the taxonomies of learning outcomes, provide insight into the assessment tools that can be utilized by stakeholders to measure the outcomes of team learning in the workplace.

Affective Outcomes of Team Learning

The importance of affect in teams is just beginning to garner serious attention as part of a broader "affective revolution" in the behavioral sciences (Ashkanasy, 2004). Theories of team performance and taxonomies of learning outcomes suggest affective states such as team efficacy, potency, and cohesion are vital to cultivating effectiveness. These affective states emerge as a result of team learning and can serve to drive team performance. For example, team efficacy (Lindsley, Brass, & Thomas, 1995), or collective efficacy when the team is the referent (Bandura, 1997), has been defined as a shared belief that a team can successfully execute its specific tasks. In contrast to task specific beliefs, potency has been defined as a set of generalized beliefs about a team's capability to execute across varying tasks and contexts (Guzzo, Yost, Campbell, & Shea, 1993). Cohesion has been defined as the result of the forces acting on team members to remain in the team (Festinger, 1950).

Mounting empirical research findings support the pivotal role of shared affect in teams. For example, meta-analytic findings suggest that team efficacy (P = .41) and potency (P = .37) are moderately related to team performance (Gully, Incalcaterra, Joshi, & Beaubien, 2002). Meta-analytic findings also support the relationships between the three facets of cohesion proposed by Festinger (1950) and team performance (Beal, Cohen, Burke, & McClendon, 2003). Collectively, these meta-analytic findings underscore the importance of emergent affective states to executing team performance and achieving team effectiveness.

Behavioral Outcomes of Team Learning

Team level skill-based learning outcomes include a wide range of latent competencies that are enacted as manifest performance processes during task episodes. These interdependent activities comprise the core of teamwork; and as teams recursively display these behavioral

processes and draw upon and update shared emergent cognitive states, team performance compiles over time (Stagl, Salas, & Burke, 2006). Teams can learn to more effectively coordinate member exchanges, distribute their workload, solve social problems, and make decisions (Campbell & Kuncell, 2002). Teams can also learn to construct higher quality solutions to challenges (Marshall-Mies, Fleishman, Martin, Zaccaro, Baughman, & McGee, 2000), formulate and execute plans more effectively (Burke et al., in press), regulate their attention and effort (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004), collaborate to process information (Ellis & Bell, 2005), and enhance their leadership capacity (Day, Gronn, & Salas, 2004). These team skills are acquired, compiled, automatized, and tuned (Kraiger, 2002). For example, team leaders provide expert coaching from which teams learn to coordinate their actions (Burke, Stagl, Klein, Goodwin, Salas, & Halpin, 2006; Stagl, Salas, & Burke, in press).

Cognitive Outcomes of Team Learning

The third class of outcomes that result from team learning are cognitive in nature. Teams can acquire declarative knowledge (i.e., information about what), procedural knowledge (i.e., information about how), and strategic or tacit knowledge (i.e., information about why and when) (Anderson, 1985). When this knowledge becomes shared via sensemaking and communication, teams can develop clusters of knowledge structures such as team mental models and transactive memory systems. Team mental models refer to team member knowledge that is either shared (Cannon-Bowers, Salas, & Converse, 1993) or compatible (Kozlowski, Gully, McHugh, Salas, & Cannon-Bowers, 1996). Conversely, transactive memory systems refer to knowledge of the distribution of information within teams (i.e., who knows what) (Wegner, 1995).

These cognitive structures provide a common conceptual framework which enables teams to perceive, represent, interpret, and respond to dynamic environments in a synchronized,

adaptive fashion (Endsley & Pearce, 2001). Teams acquire these cognitive reservoirs from their ongoing experiences. Of note, however, shared cognitive states can also be developed via a leader's efforts to synchronize task and learning cycles (Kozlowski et al., 1996) and by providing verbal briefings (Marks, Zaccaro, & Mathieu, 2000; Smith-Jentsch, Zeisig, Acton, & McPherson, 1998). Research evidence also suggests that team mental models can be developed via team interaction training (Marks et al., 2000), cross training (Marks, Sabella, Burke, & Zaccaro, 2002), and team self-correction training (Blickensderfer, Cannon-Bowers, & Salas, 1998).

Assessment of Team Learning

The previous section introduced the *ABCs of Learning*, highlighting some of the affective, behavioral, and cognitive outcomes that are often measured as proxies of team learning. This section introduces basic concepts in measurement in order to set the stage to advance a set of assessment tools that can be utilized to scale the outcomes of team learning in the workplace. Although the discussion is predominantly concerned with assessment tools that can be utilized to scale the proximal effects of learning (e.g., questionnaires, paired comparisons), some of the techniques (e.g., behavioral anchored rating scales) advanced below can also be designed and utilized to capture the actual behavioral activities that are indicative of the team learning construct itself. When feasible, practitioners are advised to measure both learning activities and the outcomes of learning in order to better gauge team development.

Assessments centers were excluded from this chapter as a means of assessing the outcomes of team learning, even though the diagnostic devices most often associated with them are described (e.g., performance tests). Assessment centers, whether conceived of as a means of facilitating holistic assessment, or in a more modern elementalistic form, are not advanced in this chapter as a means of scaling team learning because they are more appropriately conceptualized

as a process that consists of multiple assessment tools (Burroughs & Skelton, 1989). Rather than a stand alone assessment tool, an assessment center is "a comprehensive, standardized procedure in which multiple assessment techniques such as situational exercises or job simulations are used to evaluate individual employees for various purposes" (Thornton, 1992, p.1).

Measurement & Scaling

Determining whether a team has developed affective, behavioral, and cognitive outcomes as a result of its learning activities is a matter of measurement. As noted by Nunnally and Bernstein (1994, p.3), "…measurement consists of rules for assigning symbols to objects so as to (1) represent quantities of attributes numerically (scaling) or (2) define whether the objects fall in the same or different categories with respect to a given attribute (classification)." The concern in this chapter is with scaling the learning outcomes of teams. Irregardless of whether classical or modern psychometric theory is applied to scale team learning outcomes, team members, or other stakeholders with pertinent information about the team, are asked to provide their judgments (i.e., when there is a correct response) or sentiments (i.e., when measuring attitudes, values). Member responses are then combined in some manner to reflect team level outcomes.

The major requirement in scaling team outcomes is that alternative scalings be monotonically related to one another, such that different scalings rank order teams in the same way. In order to meet this requirement, practitioners are advised to develop a scaling model or internally consistent plan that specifies how to construct and apply an assessment tool. A scaling model makes testable predictions about a set of observations that are used to generate one or more dimensions (i.e., cohesion, psychological safety) to locate a team. Specifically, a scaling model establishes a hypothesis that suggests if a given scale is working, subsequent data should

support it. If accumulated observations fit a scaling model (i.e., transitivity), the model accurately determines the scale value that should be assigned to each observation.

There are four types of psychometric models that can be used to scale people and thereby the learning outcomes of teams including: nonmonotone deterministic models, nonmonotone probabilistic models, monotone probabilistic models with specified distribution forms, and monotone probabilistic models with unspecified distribution forms (Nunnally & Bernstein, 1994). In practice, the first two types of models are utilized infrequently. Moreover, while monotone probabilistic models which assume ogival item trace lines or item characteristic curves serve as the basis for modern psychometric theory (e.g., item response theory), they are also seldom found in the workplace. Therefore, these first three types of models are only briefly introduced below and the emphasis in this chapter is placed on the use of monotone probabilistic models with unspecified distribution forms for scaling the learning outcomes of teams.

Nonmonotone deterministic models are characterized by trace lines that are step functions because they assume there is no error. For example, Guttman scales which consist of dichotomously scored items assume that if a person passes an item with a given difficulty, then that person will pass all easier items on the measure. Likewise, if a person fails an item of a given difficulty, then that person will fail all of the more difficult items presented. Similarly, these models assume that if an item is passed by a person with a given level of ability, then that item will also be passed by all other individuals with higher ability. Unfortunately, infinite discrimination at the threshold is as assumption that is rooted in psychophysics and is much less realistic under the conditions in which measurement is typically conducted in the workplace.

Nonmonotone probabilistic models are similar to nonmonotone deterministic models in that both types of models assume that trace lines change slope from positive to negative or from

negative to positive. In contrast to deterministic models, however, all probabilistic models assume that trace lines are not step functions (Nunnally & Bernstein, 1994). Specifically, nonmonotone probabilistic models assume item curves in the shape of the normal distribution. These models have been used to develop Thurstone scales of attitudes. In general, however, nonmonotone items are difficult to construct as they are often characterized by double barreled item stems (Nunnally & Bernstein). Moreover, in particular, Thurstone scales are tricky to design because raters often have difficulty distancing themselves from their own beliefs and attitudes when rating the favorability of these items during development (Nunnally & Bernstein).

Monotone probabilistic models with specified distribution forms make assumptions about the fit of item trace lines to specific statistical functions (Nunnally & Bernstein, 1994). As noted above, those models that assume ogival trace lines are increasingly prevalent as they provide the basis for modern psychometrics. Perhaps the defining characteristic of most of these models is their implication for calculating scale scores. When using monotone probabilistic models with ogive item characteristic curves scale scores are calculated from the pattern of responses instead of the sum of the number of items answered correctly (i.e., in the alpha direction). Despite the convenient mathematical properties that characterize these models, the conventional approach of summing items answered in the alpha direction remains the simplest model for scaling people or the learning outcomes of teams, as discussed next (Nunnally & Bernstein).

As noted, monotone probabilistic models with unspecified distribution forms such as the linear model are the most popular means of constructing assessment tools. The linear, centroid, or summative model is so labeled because scale scores are a function of the linear combination or sum of item responses (Nunnally & Bernstein, 1994). Given monotonic item trace lines and a unitary attribute as the focus of measurement, the linear model proves versatile, as it can be used

to guide the construction of weighted and unweighted multicategory and dichotomous items. The intuitive appeal and widespread applicability of the linear model makes for a compelling case to utilize it when scaling the learning outcomes of teams. For example, the linear model can be used as a basis for constructing Likert scales to measure shared affect (e.g., cohesion), or when designing behaviorally anchored rating scales for measuring team processes (e.g. coordination). The linear model underpins the multi-item assessment tools that are discussed in this chapter.

Measuring Shared Affect

One set outcomes that arise from the learning activities teams engage in are affective in nature. Shared affective states such as trust, collective orientation, potency, collective efficacy, and cohesion originate in the individual level affect of team members and via communication, shared experience, and dyadic interactions manifest as collective phenomena which can be scaled as outcomes of team learning. In practice, affective outcomes are often measured via the use of inventories and questionnaires that require respondents to make ratings about their (dis)agreement to a set of statements presented on a Likert scale (Gregorich, Helmreich, & Wilhelm, 1990). The items of these scales often reference the team as the unit of interest, and given sufficient perceptual agreement individual responses are aggregated to the team level. Conversely, a consensus approach involves a single response to an item from the entire team which has been determined via joint discussion. Although a consensus approach is a less common approach to measuring affective states (Gully et al., 2002), evidence suggests that consensus ratings can explain significantly greater variance in team effectiveness than do aggregate indices (Kirkman, Tesluk, & Rosen, 2001). The widespread use of inventories and questionnaires suggests some additional commentary on these tools may be valuable.

The use of self-report measures can be traced back 150 years to sir Francis Galton's investigations of mental imagery. Galton believed that character, conceptualized by him as a respondent's inner world of perception and emotion, was a definite and durable phenomenon that could be measured. To measure character, Galton developed standardized questionnaires. A measure is standardized to the extent that it has clear rules for assigning numbers to attributes, is practical to apply, does not demand extensive skill to administer, and produces results that are not administrator specific (Nunnally & Bernstein, 1994). There are many advantages to standardized assessment, as it: facilitates observation, replication, and generalization, provides finer distinctions than those afforded by subjective judgments, permits the use of more powerful analytics, and provides an economical means of assessing large numbers of respondents (Nunnally & Bernstein). Today, self-report inventories and questionnaires are the most widely utilized form of assessment (Hough & Ones, 2001). In fact, their abundance prompted Cattell (1963) to characterize these diagnostic devices as the 'rabbits' of the psychometric world.

There are a variety of inventories, questionnaires, and surveys (e.g., checklists, scaled response, multiple choice / forced choice) that can be utilized to assess attitudes (Guion, 1998). For example, checklists are comprised of a list of words or phrases that respondents check off as applicable. Scaled response inventories are characterized by three or more options that are arranged in an ordered sequence. Many of the instruments utilized to scale shared affective states include a five point scale to which respondents rate whether they '*strongly disagree*', '*disagree'*, '*can not say'*, '*agree'*, or '*strongly agree'* that a given item stem is descriptive of their own internal states or those of their team. A third type of inventory utilizes a multiple / forced choice approach. A forced choice approach is characterized by multiple response options from which respondents must chose which stimuli is most descriptive and which is least descriptive.

Although self-report inventories and questionnaires provide a number of advantages, their use comes with a few notable caveats. For instance, the use of self-report measures assume that respondents have the awareness to accurately assess their own, or their team's attributes, a sometimes questionable assumption. Inventories and questionnaires are also subject to scrutiny because of an inherent vulnerability to contrived response protocols. A response set is a tendency to adopt a certain style in responding to stimuli (Guion, 1998). There are a number of possible response sets that self-report inventories and questionnaires are susceptible to including socially desirable responding and acquiescence (Kurtz & Parrish, 2001). This is an important issue, because as noted by Guion (1998, p. 596) "A response set contributing variance to scores intended for inferences about something else is a validity-reducing contaminant".

A social desirability response set is characterized by intentional distortion (i.e., faking) on the part of respondents in order to look good to the administrators or other stakeholders making decisions based in part on the information provided by the administration of a given measure. Examining this issue, Messick (1960) identified nine distinct social desirability factors apparent in one commonly utilized self-report instrument. Self-report scales are particularly subject to social desirability response sets because they are often comprised of items that are translucent to respondents in regards to which option is organizationally valued. Therefore, those respondents adopting a social desirability response set will, when encountering the item stem ' How confident do you feel that your team is able to recognize complex problems', be more to answer *'extremely confident'* rather than admitting that they are '*somewhat confident*'. Moreover, faking will be more likely to occur when respondents hold the expectation that intentional distortion is instrumental in achieving valued outcomes, as would be the case when a team is assessed at the

conclusion of certification training conducted to shape a team's attitudes about provision of exemplary customer service. These issues warrant consideration when measuring shared affect.

Measuring Shared Skills

The second set of outcomes that compile from team learning are shared behavioral and cognitive skills. Latent team skills are enacted episodically as performance processes during the transformation of raw materials to goods and services. Teams can learn to more effectively coordinate their exchanges, cooperate when dynamically allocating their resources, and communicate to make better decisions as they navigate their internal and external challenges (Campbell & Kuncell, 2002; Kozlowski & Bell, 2003; Salas, Stagl, & Burke, 2004). Team skills that result from learning are often assessed during performance tests which typically include supervisor, peer, subject matter expert, or administrator ratings. Of note, however, one type of performance test, situational judgment tests, require respondents to identify what they should or would do in response to a posed scenario rather than actually acting out a chosen response.

Performance tests encapsulate a wide range of stimuli that: (1) can be cognitive or noncognitive in nature, (2) require paper and pencil or hands on responses, and (3) utilize a forced choice or constructed response format (Guion, 1998). Even though there are a wide range of performance tests, all of these devices generally simulate key aspects of a work situation such as contextual forces and complex problems to be solved. Performance tests are typically utilized to measure proficiency, skill, or understanding. Although prediction is implied, the emphasis is placed on current rather than future performance (Cascio, 1998). Therefore, these assessment tools are most often associated with capturing samples rather than signs of behavior.

Historically speaking, performance tests were one the earliest measurement strategies utilized to assess skills, as seen by Münsterberg's (1912) use of these tools to select trolley car

operators almost a century ago. Not long after, Max Simoneit utilized performance tests to help the German army select officers during WWII; a practice that was quickly benchmarked by both Britain's War Office Selection Board and in the United States by the Office of Strategic Services (OSS) (Highhouse, 2002). For example, the OSS utilized the Construction Test to select spies and saboteurs. This exercise required participants to persuade two passively hostile subordinate confederates to complete a project (i.e., assembling a shelter from poles) in a brief period of time. The two confederate subordinates made every effort to loaf or blatantly sabotage the project, and even went so far as to purposely wreck the structure after its completion.

The approach adopted by the OSS, and other governmental organizations, exemplified how situational exercises could be designed to prompt individual assesses, or even entire teams, to display their skills. Once enacted, assessors would form overall impressions about a candidate's capabilities. Specifically, assessors captured relevant information, recorded it in some manner, and recalled it later to form an impression about a candidate. Unfortunately, the process of attention, encoding, recalling, and integrating information in order to form judgments is susceptible to a number of random and systematic errors (Guion, 1998). For example, ratings are subject to distributional errors in mean levels (e.g., leniency, severity) and variance (e.g., restriction of range). Furthermore, ratings are subject to a host of other errors such as illusory halo and similar to me error. In retrospect, the approach employed in these early assessment centers was holistic by design and afforded little attention to psychometric concerns. The assessment process was characterized by: (1) an active orientation whereby assesses acted out their responses to contrived scenarios, (2) a limited concern for psychometric issues as instruments were often modified on the fly to test a particular assessor's hunches about a given

candidate, and (3) a scoring process that was heavily dependent upon the idiosyncrasies inherent to the intuitive clinical judgment of a given assessor or assessment team (Highhouse, 2002).

By the 1960's, personnel psychologists had successfully lobbied both the public and private sectors of the value of testing in the workplace. It was during this historical period that the military's initial approach to performance assessment via the use of situational tests was revamped by government scientists and their private sector colleagues. This massive effort was undertaken to address some of the serious psychometric concerns raised about performance testing as typified by the holistic assessment process utilized by the OSS. Modifications made to situational tests emphasized standardization of testing conditions and scoring procedures (Flanagan, 1954). Rater errors were reduced via the use of error resistant forms, rater error training, and frame of reference training. For example, assessor ratings were enhanced via the use of checklists of behavior to make performance ratings. Moreover, the use of graphic rating scales (Meister, 1985), behavioral anchored rating scales (Smith & Kendall, 1963), and behavioral observation scales (Tziner, Joanis, & Murphy, 2000), helped reduce errors because they served to minimize assessor idiosyncrasies, thereby making ratings more standardized and increasing cross ratee comparability. Furthermore, these new rating scale formats were easy to understand and use in fluid settings and thereby contributed to their wide spread acceptance, frequent application, and timely completion.

It was due in part to this overhaul that performance tests became ubiquitous with assessment centers (ACs) conducted in the private sector. Widely acclaimed ACs that included performance tests were conducted at AT&T (see Bray & Grant, 1966) and Standard Oil of New Jersey (see Sparks, 1990). Similar to their military predecessors, these interventions also required elaborate facilities and rigorous administrator training (Anastasi, 1988). Perhaps the parallels are

not surprising given that Douglas Bray, a graduate student working with the OSS assessor team, was also a key figure in the use of ACs at AT&T. Irregardless of the similarities and differences between the public and private sector uses of performance testing, the results from AT&T, Standard Oil (i.e., Exxon), and other notable interventions, were promising, and performance testing became widely applied in talent assessment interventions conducted to assess skills.

Today, many governments and their affiliated industries utilize performance tests to assess teams as part of training, development, and certification initiatives (e.g., Baker, Prince, Shrestha, Oser & Salas, 1993; Bell, 1995; Bowers, Salas, Prince & Brannick, 1992; Jentsch & Bowers, 1998; Smith-Jentsch, Salas & Baker, 1996). Although performance tests have proven most useful for assessing team skills which have compiled from learning activities, research has continued to examine a wide range of issues inherent to the application of these assessment tools. For example, research has been undertaken to investigate the optimal fidelity required of performance tests (Marks, 2000). The evidence accumulated from this research suggests that the physical fidelity of an environment created within a simulation is of secondary concern relative to psychological reality or fidelity. Performance tests that simulate operational settings should be designed to trigger similar psychological processes as those experienced in the work context. Specifically, Marks' research suggests that the fidelity of a performance test should match the level of complexity typically experienced by an individual or team during their performance episodes. Thus, in some circumstances, the use of high fidelity performance tests to provide an immersive experience may be warranted. Of note, however, as workers increasingly interact via email and voicemail, lower fidelity simulations that only require written or oral responses may sufficiently mirror the cognitive processes that actually occur during work. The findings of this

line of research suggest that lower fidelity pc based performance tests may be a cost effective and psychometrically sound solution for organizations seeking to assess team skills.

A lower fidelity alternative to hands on performance tests for assessing team skills is offered by situational judgment tests (SJTs) (see Weekley & Ployhart, 2006). SJTs are lower fidelity assessment tools that are comprised of item stems, response instructions, standardization rules, and when a forced choice format is used, a set of response alternatives. SJTs can be administered via paper-and-pencil, video, or orally. According to Fritzsche, Stagl, Salas, and Burke (2006, p.301), the use of SJTs is predicated upon the assumption that those respondents:

> …who can identify more effective and less effective responses to job-related situations have greater job-relevant knowledge or better job-related judgment and reasoning skills and are thus, expected to have higher levels of job performance than those who are less able to identify appropriate responses to job-related situations.

SJTs are quite versatile, as they can be utilized to assess a wide range of team skills that result from team learning, including: crew resource management skills (Hedge, Borman, & Hanson, 1996), judgment skills (Hunter, 2003), problem solving skills (Motowidlo, Dunnette & Carter, 1990), decision making skills (Stagl, Fritzsche, Salas, Wooten, & Burke, under review), and interpersonal skills (Olson-Buchanan, Drasgow, Moberg, Mead, & Donovan, 1998). One example of a paper-and-pencil-based SJT is the Tacit Knowledge for Military Leadership (TKML) (Sternberg, Forsythe, Hedlund, Horvath, Wagner, Williams, Snook & Grigorenko, 2000). Sternberg and colleague's developed the TKML as a low fidelity simulation to measure the tacit knowledge underlying practical intelligence. Although the TKML was designed to measure knowledge, rather than skill per se, Sternberg and colleagues' suggest that SJTs designed with an extended case study format can be utilized to assess the specific cognitive and

decision making skills inherent to problem formulation and solution generation (Cianciolo, Antonakis & Sternberg, in press). An extended case study exercise incorporates a (1) detailed overview, (2) additional documentation, and (3) a constructed response format. Thus, an SJT with an extended case study format may provide and more cost effective means of assessing team skills than more expensive and time consuming hands on performance tests.

Measuring Team Knowledge

The third set of team level outcomes that emanate from team learning are shared cognitive structures such as team mental models. Although team mental models manifest in a variety of specific forms (e.g., task, equipment, team member, team interaction), each of these models serves the similar purpose of facilitating the organization and acquisition of information in a task domain, which, in turn, helps team members predict the need for and subsequently execute coordinated actions (Cannon-Bowers et al., 1993). Thus, team knowledge provides an infrastructure for enacting episodic team processes (Mathieu & Schulze, in press).

The assessment of team mental models is concerned with both the degree of convergence amongst team members on the components of a given model and the relationships between these components (Mohammed, Klimoski, Rentsch, 2000). In regards to assessment, the first of these two issues is concerned elicitation, or techniques that are utilized to determine the content of a team mental model (Klimoski & Mohammed, 1994). The second issue is concerned with representation, or tools that can be utilized to illuminate the relationships between components. This subsection includes a discussion of pairwise relatedness ratings as an elicitation technique and the use of Pathfinder and UNICET as tools for representing data obtained from these similarity ratings. We also discuss the use of card sorting and concept mapping as approaches that combine elicitation and representation.

The underlying assumption of pairwise relatedness ratings for scaling team learning outcomes is that the closer raters perceive two concepts to be, the closer those concepts are in their knowledge structures (Kraiger et al., 1993). In regards to this issue, Thurstone's Law of Comparative Judgments provides a basis for deriving an interval level index of this ordinal scaling approach. The process consists of presenting each rater with all possible pairs of stimuli. Raters are then asked whether stimulus A is similar to stimulus B on some attribute in question. Observed rating are then analyzed to determine the proportion of times stimulus A is deemed similar to stimulus B etc. and these proportions are converted to z scores to form a matrix of normal deviate units. Matrix entries are then summed and averaged for each column in order to reduce measurement error. The lowest average is subtracted from itself and the remaining averages to obtain an interval level scaling of ordinal data obtained from paired comparisons.

In practice, relatedness ratings are often post-processed via the use of statistical software programs such as Pathfinder Associate Networks (Schvaneveldt, 1990). For example, the Pathfinder algorithm has been utilized to scale the knowledge structure coherence of trainees exposed to mastery or performance training goals (see Kozlowski, Gully, Brown, Salas, Smith, & Nason, 2001). Pathfinder converts paired comparisons to nodes (i.e., concepts) and node linkages (i.e., relations between concepts) in order to assess cognitive structure. A related algorithm, UNICET, has also been utilized to scale team mental models (see Mathieu, Heffner, Goodwin, Cannon-Bowers, & Salas, 2005; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). UNICET provides indices of both centrality and convergence of mental models via the use of QAP correlations between two matrices (Borgatti, Everett, & Freeman, 1992).

Two other techniques for scaling team mental models are card sorting and concept mapping. The card sorting approach requires raters to cluster concept cards together based upon

their perceived similarities and differences. When concept mapping is used, respondents are asked to specify the relationships between concepts. Concepts maps are comprised of the previously described nodes and links and one additional component, propositions, which describe the meaning of the linkage between two nodes. A recently developed software tool, TPL-KATS, can be utilized to automatically administer and score both card sorts and concept maps (Hoeft, Jentsch, Harper, Evans, Berry, Bowers, & Salas, 2002). TPL-KATS thus provides a relatively quick and cost efficient means of scaling a team's mental models, as well as comparing these knowledge structures to an expert's models.

Conclusion

The use of teams as a preferred performance arrangement is increasingly prevalent in the workplace. Teams are a popular means of structuring work, in part, because their members can lean on and learn from one another during performance episodes. In fact, an accumulating base of research findings suggest learning is vital to both routine and adaptive team performance. This chapter leveraged some of the theories, measurement techniques, and lessons learned in the behavioral sciences to advance a set of assessment tools that can be used to measure learning and its outcomes in teams. We hope that the advanced set of assessment tools will help practitioners measure growth and decline in 'dynamic learning teams'.

Acknowledgement

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported in part by funding from the Army Research Laboratory's Advanced Decision Architecture Collaborative Technology Alliance (Cooperative Agreement DAAD19-01-2-0009). This work was also supported in part by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

References

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.

Anderson, J. P. (1985). *Cognitive psychology and its implications*. (2nd ed.). New York: Freeman.

Argyris, C. & Schon, D. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addision-Wesley.

Baker, D. P., Prince, C., Shrestha, L., Oser, R. & Salas, E. (1993). Aviation computer games for crew resource management training. *The International Journal of Aviation Psychology, 3*, 143-156.

Bell, H. H. (1995). The engineering of a training network. In Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting (pp. 1311-1315). Santa Monica, CA: Human Factors and Ergonomics Society.

Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (1998). Cross-training and team performance. In J. A. Cannon-Bowers, & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 299-311). Washington, DC: APA.

Bowers, C. A., Salas, E., Prince, C., & Brannick, M. (1992). Games teams play: A method for investigating team coordination and performance. *Behavior Research Methods, Instruments, & Computers, 24,* 503-506.

Burke, C. S., Stagl, K. C., Klein, C., Goodwin, G. F., Salas, E., & Halpin, S. (in press). What types of leader behaviors are functional in teams?: A meta-analytic integration. *Leadership Quarterly*.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (in press). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology.*

Burroughs, W. & Skelton, A. (1989). Assessment centers yield unbiased employee evaluations. *Staffing and Recruitment, 1*, 33-39.

Campbell, J. P. & Kuncel, N. R. (2002). Individual and team training. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (pp. 272-312). London, England: Sage.

Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology, 46*, 823-850.

Cannon-Bowers, Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan Jr. (Ed.), Current *issues in individual and group decision making* (pp. 221-246). Hillsdale, NJ: Erlbaum.

Cascio, W. F. (1998). *Applied psychology in human resource management*. Prentice Hall: Upper Saddle River, NJ.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.

Cianciolo, A. T., Antonakis, J. & Sternberg, R. J. (2004). Practical intelligence and leadership: Using experience as a mentor. In D. Day, S. J. Zaccaro, & S. M. Halpin (Eds.), *Leader development for transforming organizations*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Day, D. V., Gronn, P., & Salas, E. (2004). Leadership capacity in teams. *The Leadership Quarterly, 15*, 857-880.

Day, D. & Lance, C. (2004). Understanding the development of leadership complexity through latent growth modeling. In D. Day, S. J. Zaccaro, & S. M. Halpin (Eds.), *Leader development for transforming organizations*. Mahwah, NJ: Erlbaum Associates, Inc.

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (in press). A multiple goal, multilevel model of feedback effects on the regulation of individual and team performance in training. *Journal of Applied Psychology*.

Dickinson, T.L. & McIntyre, R.M. (1997). A conceptual framework for team measurement. In M.T. Brannick, E. Salas, & C. Prince (Eds.), T*eam Performance and Measurement: Theory, Methods, and Applications.* (pp.19-43). Mahwah, NJ: Erlbaum Associates.

Edmondson, A.C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44(2),* 350-383.

Edmondson, A.C., Bohmer, R.M., & Pisano, G.P. (2001). Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative Science Quarterly, 46(4),* 685-716.

Ellis, A.P.J., Hollenbeck, J.R., Ilgen, D.R., Porter, C.O., West, B.J., & Moon, H. (2003). Team learning: Collectively connecting the dots. *Journal of Applied Psychology, 88(5),* 821-835.

Fritzsche, B. A., Stagl, K. C., Salas, E. & Burke, C. S. (2006). Enhancing the design, delivery and evaluation of scenario-based training: Can situational judgment tests contribute? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests*. Mahwah, NJ: Lawrence Erlbaum Associates.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum & Associates.

Gully, S. M, Incalcaterra, K. A., Joshi A., & Beaubien, J. M. (2002). A meta-analysis of team efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships. *Journal of Applied Psychology, 87,* 819-832.

Hackman, J. R. (1987). The design of work teams. In J. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315-342). Englewood Cliffs, NJ: Prentice-Hall.

Hedge, J. W., Borman, W. C. & Hanson, M. A. (1996, March). *Videotaped crew resource management scenarios for selection and training applications*. Paper presented at the 38th annual conference of the International Military Testing Association.

Highhouse, S. (2002). Assessing the candidate as a whole: A historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology, 55,* 363-387.

Hough, L. M. & Ones, D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.). *Handbook of industrial, work and organizational psychology*. Thousand Oaks, CA: Sage Publications Inc.

Hunter, D. R. (2003). Measuring general aviation pilot judgment using a situational judgment technique. *The International Journal of Aviation Psychology, 13,* 373-386.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From Input-Process-Output models to IMOI models. *Annual Review of Psychology, 56,* 1-19.

Jentsch, F. & Bowers, C. A. (1998). Evidence for the validity of PC-based simulations in studying aircrew coordination. *International Journal of Aviation Psychology, 8*, 261-276.

Klimoski, R., & Mohammad, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20*, 403-437.

Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 333-375). London: Wiley.

Kozlowski, S.W.J., Gully, S.M., McHugh, P.P., Salas, E., & Cannon-Bowers, J.A. (1996). A dynamic theory of leadership and team effectiveness: Developmental and task contingent leader roles. In G.R. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 14, pp. 253-305). JAI Press.

Kozlowski, S. W. J., Gully, S. M., Salas, E., & Cannon-Bowers, J. A. (1996). Team leadership and development: Theory, principles, and guidelines for training leaders and teams. In M. Beyerlein, S. Beyerlein, & D. Johnson (Eds.), *Advances in interdisciplinary studies of work teams: Team leadership* (Vol.3, pp. 253-292). Greenwich, CT: JAI.

Kozlowski, S. W. J. & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco, CA: Jossey-Bass.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.

Kurtz, J. E. & Parrish, C. L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R. *Journal of Personality Assessment, 76*, 315-332.

Levitt, B. & March, J.G. (1988). Organizational learning. *Annual Review of Sociology, 14*, 319-340.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team process. *Academy of Management Review, 26*, 356-376.

Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology, 87*, 3-13.

Marks, M. A., Zaccaro, S. J., & Mathieu, J. E. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *Journal of Applied Psychology, 85*, 971-986.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*, 273-283.

Mathieu, J. E. & Schulze, W. (in press). The influence of team knowledge and formal plans on episodic team process → performance relationships. *Academy of Management Journal*.

Messick, S. (1960). Dimensions of social desirability. *Journal of Consulting Psychology, 24*, 279-287.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.

Nunnally, J. & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw Hill.

Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology, 51*, 1-24.

Salas, E., Stagl, K. C. & Burke, C. S. (2004). 25 years of team effectiveness in organizations: Research themes and emerging needs. In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology*. New York: John Wiley & Sons.

Salas, E., Stagl, K. C., Burke, C. S. & Goodwin, G. F. (2006). Fostering team effectiveness in organizations: Toward an integrative theoretical framework of team performance. In J. W. Shuart, W. Spaulding, & J. Poland, (Eds.), *Modeling complex systems: Nebraska Symposium on Motivation, 52,* Lincoln, NE: University of Nebraska Press.

Smith-Jentsch, K., Salas, E., & Baker, D. P. (1996). Training team performance-related assertiveness. *Personnel Psychology, 49*, 909-936.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds.) *Making decisions under stress* (pp. 271-297). Washington, DC: APA.

Stagl, K. C., Fritzsche, B. A., Salas, E., Wooten, W., & Burke, C. S. (under review). The construct validity of a situational judgment test in a maximum performance context. *International Journal of Selection & Assessment*.

Stagl, K. C., Salas, E., & Burke, C. S. (in press). Best practices in team leadership: What team leaders do to facilitate team effectiveness. In J. A. Conger & R. E. Riggio (Eds.), *The practice of leadership* (pp.). John Wiley & Sons.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., Snook, S. A. & Grigorenko, E. L. (2000). Practical intelligence: An example from the military. *Practical intelligence in everyday life*. Cambridge, UK: Cambridge Press.

Tannenbaum, S. I., Beard, R. L., & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. In K. Kelley (Ed.), *Issue, theory, and research in industrial/organizational psychology* (pp. 117-153). Amsterdam: Elsevier

Thornton, G. C. (1992). *Assessment centers is human resource management*. Reading MA: Addison-Wesley.

Van Der Vegt, G.S., & Bunderson, J.S. (2005). Learning and performance in multidisciplinary teams: The importance of collective team identification. *Academy of Management Journal*, *48(3)*, 532-547.

Weekley, J. A. & Ployhart, R.E. (2006). (Eds.), *Situational judgment tests*. Mahwah, NJ: Lawrence Erlbaum & Associates.

Wenger, D. M. (1995). A computer network model of transactive memory. *Social Cognition, 13*, 319-339.

Running head: Cross Training Teams

Best Practices in Cross Training Teams

Kevin C. Stagl, Eduardo Salas, & Stephen M. Fiore

Department of Psychology

&

Institute for Simulation & Training

University of Central Florida

DO NOT CITE OR DISTRIBUTE WITHOUT PERMISSION OF THE AUTHORS

Address correspondence to: Kevin C. Stagl
Institute for Simulation & Training
University of Central Florida
3100 Technology Parkway
Orlando, FL 32826
(407) 739-0717 (w); 658-5059 (fax)
kstagl@ist.ucf.edu

Stagl, K. C., Salas, E., & Fiore, S. M. (in press). Best practices in cross training teams. In D. A. Nembhard (Ed.), Workforce cross training handbook (pp.). Boca Raton, FL: CRC Press.

Best Practices in Cross Training Teams

The milieu in which an increasing number of businesses, alliances, and government brokered coalitions operate is typified by hyper competitive global markets, rapidly evolving ambiguous situations, imperfect solutions, information overload, intense time pressure, and where there are severe consequences of error (Orasanu & Salas, 1993). In response to the difficult tasks and complex challenges inherent to today's workplace, organizations often use teams as a performance arrangement to structure work. Teams thrive in this controlled chaos because they can leverage superior resources when identifying emerging opportunities and formulating and executing plans accordingly. In fact, effective team performance can result in improved: financial performance, adaptation, efficiency, productivity, safety, service quality, satisfaction, and commitment.

While teams can yield an impressive array of results, the ground truth is teams of experts sometimes fail with disastrous consequences for those directly involved and indirectly vested. Tragic team failures abound in the public (e.g., the Federal Emergency Management Agency's leadership team's response to Hurricane Katrina) and private sector (e.g., the 160 fatalities attributed to crew teamwork failures aboard American Airlines flight #965). Why do teams of well intentioned, highly motivated, seasoned professionals fail on the grandest stages and what can be done about it? Clearly, the answers to these questions are complex but insight is afforded by both the science of teams and the science of team training (Salas & Cannon-Bowers, 2001). The principles, guidelines, and lessons learned in these domains can be applied to transform a team of experts into an expert team.

One notion that is useful for understanding both exceptional and ineffective team

performance is shared mental models. Shared mental models are emergent cognitive states that consist of commonly held organized knowledge structures. These dynamically activated memory structures are representations of reality that allow team members to anticipate the need for, and select actions that are coordinated with, those of their teammates (Cannon-Bowers, Salas, & Converse, 1991; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). Specifically, shared mental models are drawn upon by members before, during, and after they enact the taskwork and teamwork processes that comprise both routine and adaptive team performance (Burke, Stagl, Salas, Pierce, & Kendall, in press). Thus, shared mental model theory is relevant to training teams to be more effective because it provides a basis for understanding how team members conduct the cognitive components of their jobs (Cannon-Bowers & Salas, 1990). For example, shared mental models can be used to understand and improve how members recognize, integrate, and apply information to make decisions in naturalistic settings (Klein, 1989).

One human capital management intervention that has successfully been used to foster shared mental models and thereby effective team performance is cross training (Salas, Dickinson, Converse, & Tannenbaum, 1992). Cross training is an instructional strategy that consists of a suite of interrelated techniques (e.g., positional clarification, positional modeling, positional rotation) which are used to reduce interpositional uncertainty by imparting interpositional knowledge (IPK) (Baker, 1991). IPK is a specific type of shared team-interaction mental model that consists of role knowledge of one's teammates' task requirements, functions, information needs, action-outcome contingencies, and coordination demands (Fiore, Johnston, & McDaniel, 2005; Kahn, Wolfe, Quinn, Snoek, & Rosenthal, 1964). IPK also includes contextually specific

knowledge about temporal contingencies and cause-and-effect task associations (Volpe, Cannon-Bowers, Salas, & Spector, 1996). By fostering IPK, cross training is a means of developing emergent cognitive states such as shared mental models that, in turn, facilitate a team members' efforts to predict, explain, anticipate, and coordinate within teams. Moreover, because cross training exposes members to a variety of tasks, it can also foster intrinsically motivated, cohesive teams. These cognitive and affective states, in turn, enable the explicit and implicit coordination at the heart of team performance.

While the majority of scientific and anecdotal evidence suggests cross training is a viable means of fostering effective teams, to date, there is no comprehensive source for best practices in utilizing this instructional strategy to train teams and their members. In an effort to remedy this state of affairs, this chapter briefly reviews the nature of teams and team training in order to set the stage for a more in depth discussion of the nature of cross training. The theoretical and empirical evidence generated from cross training interventions conducted in a variety of settings is then reviewed. The findings gleaned from these interventions are summarized in a heuristic framework illustrating the proximal and distal effects of cross training. Moreover, lessons learned are distilled from these initiatives and subsequently framed as 20 best practices for cross training teams.

A Science of Teams

Teams have a long and storied history (Salas, Priest, Stagl, Sims, & Burke, 2006). In fact, the success of teams has been integral to the survival of human kind ever since early tribes used hunting parties to stalk and bag prey. Yet, it was not until the late twentieth century that theoretical and empirical work on teams began to flourish. Since

that time, research has continued unabated in a number of disciplines and from a variety of perspectives, with the common goal of advancing the science of teams.

Over the course of past of century, much has been learned about the nature of teams. An early description of teams defined them as an "interdependent collections of individuals who share responsibility for specific outcomes for their organizations" (Sundstrom, De Meuse, & Futrell, 1990, p. 120). More specifically, they were characterized as "two or more individuals who must interact and adapt to achieve specified, shared, and valued objectives" (Salas, Dickinson, Converse, & Tannenbaum, 1992, p. 4). This definition has matured to include the notion of teams as complex entities, comprised of two or more individuals, who interact socially, dynamically, episodically, and adaptively (Stagl, Salas, & Burke, in press). Team members engage in meaningful dyadic and team level exchanges before, during, and after the projects or tasks they are charged to complete. Over time, these interactions become entrained to contextual forces like supply and demand such that a pace and tempo of team performance develops (Ancona & Chong 1999). For example, the members of an Intel flex fab manufacturing team ramp up their efforts to produce components for a client's new product launch. As the half-way point to their deadline approaches, members come together with their leader to reflect upon their progress and adapt to newly arisen issues. The team then re-engages, meets its targeted goals, and ramps-down until the next set of orders exceed current production levels.

Teams also perform episodically over time. For example, a real estate development team works episodically to create valuable assets from raw land by engaging in taskwork and teamwork processes that convert inputs to outputs. These

input-process-output cycles are repeated in action and transition phases which unfold within performance episodes (Marks, Mathieu, & Zaccaro, 2001; Mathieu & Schulze, in press). The transition phase encapsulates a preliminary meeting, where the team draws upon its shared resources to discuss what it aims to accomplish and what actions are necessary to meet its objectives. During this initial phase a plan is mapped as the team considers which options are most viable and likely to reach fruition. Once a plan is formulated, the team shifts to an action phase where property is secured, developed, and marketed. After construction is finished and the property has been sold, the team reflects on its shortcomings and accomplishments and the cycle begins anew.

Another characteristic of teams is meaningful levels of task, goal, and feedback interdependencies (e.g. pooled, sequential, reciprocal, team) (Saavedra, Earley, & Van Dyne, 1993). For example, a medical team composed of surgeons, nurses, rehabilitative therapists, anesthesiologists, and other specialists work towards restoring the infirm back to health. This extended team has sequential interdependencies as each specialist contributes their expertise in turn as the patient moves through various levels of need. At the beginning of a serious injury, the individual is traumatized and surgeons work to insure survival. Afterwards, specialty nurses watch over and manage care. Eventually, physical and occupational therapists teach the injured person how to live and work productively. As each member of this extended team performs their respective duties in succession, the patient grows increasingly healthy.

Teams are also often designed with hierarchical structures to accomplish specific tasks over a limited period of time (Salas et al., 1992). These characteristics typify tactical decision-making teams such as naval gunfire-support teams, antisubmarine

warfare teams, and guided missile teams (McIntyre & Salas, 1995). For example, guided missile teams operate in a command and control structure with a formal leader and three to five team members. They interdependently seek out, share, and utilize information from their environment to make decisions about enemy engagements. These teams share the common goal of protecting the interests of the United States at home and abroad. Similarly, hierarchical structures and a limited life-span also characterize private sector management teams such as corporate executive teams and regional steering committees (Sundstrom, McIntyre, Halfhill, & Richards, 2000).

The members of teams also typically hold distinct roles. This characteristic can be seen in a football team's offensive squad, which consists of quarterbacks, running backs, receivers, tight ends, centers, guards, and tackles. Each position is typically staffed by members that play unique roles. For example, the 2005 Superbowl champion Pittsburgh Steelers of the National Football League had one running back whose role it was to play the vast majority of the team's first and second downs, a second running back that specialized in third down conversions, and a third running back who was often asked to pick up much needed short yardage particularly at the goal line. The use of each of these three roles was tightly choreographed to provide the team with the optimal mixture of capability given such factors as the down and distance on a play.

Teams also have distributed expertise. Distributed assets are apparent in jazz bands, where each musician leverages their experiences, training, and skills to make a unique contribution to a successful performance. Typically, the guitarist of a jazz ensemble does not possess the skills to play a saxophone or drums with professionalism; however, as a team they anticipate each other's actions by knowing enough about the

other's capabilities to follow chord changes and instrumental solos. Once a solo is complete, the finished musician steps back and supports the others on stage. Although they possess unique skills, they know enough about each other's capabilities to complete a song seamlessly; with emotion, expertise, and enthusiasm.

While team members often have unique responsibilities and assets, they share common valued goals. Collaborative effort towards common valued goals can be seen in the practices of Solectron Corporation, a customized electronics solutions firm based in Charlotte, North Carolina. Solectron assigns engineers to client company design teams. Solectron embeds its team members in their client's operations because they believe that 80% of a products price and supply chain performance is determined in the collaborative product design phase. In fact, their customers find that early design involvement with Solectron is crucial to the success and speed of eventual production. By rallying around the common goals of providing innovative, customized, quality electronic solutions via teamwork, Solectron and their clientele are able to truly leverage the benefits of teams.

Teams are also embedded within an organizational / environmental context that influences and is influenced by enacted competencies and processes, emergent cognitive and affective states, performance outcomes, and stakeholder judgments of team member and team effectiveness (Salas, Stagl, Burke, & Goodwin, in press). For example, the strategies and decision-making processes enacted by distributed leadership teams at IBM are shaped by their organizational and environmental context (Stagl, Salas, Rosen, Priest, Burke, Goodwin, & Johnston, in press). To capture emerging markets, IBM is moving away from a cost leader strategy emphasizing manufacturing towards an innovator strategy emphasizing the development of information technologies, B2B solutions, and

global consulting services. This shift in organizational priorities in response to a changing marketplace affects the types of strategic partners coveted by IBM, as witnessed by the recent $1.75 billion alliance with Lenovo, Asia's largest computer maker.

The handful of team characteristics expounded upon in this preliminary section is just a small sample of the myriad of attributes which typify teams. Prior research offers a more comprehensive view of the numerous factors which contribute to, comprise, flow from, and impinge upon teamwork, team performance, and perceptions of team effectiveness (see Figure 1) (Tannenbaum, Beard, & Salas, 1992). The framework of team effectiveness illustrated in Figure 1 depicts multiple inputs, throughputs, and outputs which are shaped and constrained by organizational / situational characteristics.

Tannenbaum and colleagues' (1992) framework illustrates some of the core cognitive and behavioral processes that comprise teamwork (e.g., coordination, decision making, backup behavior, mutual performance monitoring). Team members draw upon their latent resources (e.g., cognitive abilities, tacit knowledge, mental models) and teams draw upon their shared resources (e.g., shared mental models, cohesion) to execute these teamwork and taskwork processes during team performance. These actions transform raw materials and information resources into products and services. As a team performance episode unfolds, feedback is generated and team members and teams change in terms of the resources they have available for the next phase of performance.

The framework of team effectiveness advanced by Tannenbaum et al. (1992) also depicts a variety of human capital management interventions that can be leveraged by stakeholders charged with developing team member and team performance. These interventions are critical to turning a team of experts into an expert team (see Salas,

Cannon-Bowers, & Johnston, 1997). The developmental solutions highlighted by Tannenbaum and colleagues include individual training, team training, and team building. Of particular note to the current endeavor, a host of team training programs have been advanced for developing teams, one of which is cross training. A brief overview of the nature of team training is provided in the next subsection and a representative set of best practices are advanced prior to launching into a more targeted treatment of cross training.

A Science of Team Training

One means of helping to ensure the benefits of teams are reaped is to prepare them for the operational realities they will encounter via training and development initiatives. Many organizations opt for this approach, as annual domestic expenditures for workforce training and development exceed $200 billion (Carnevale, Gainer, & Villet, 1990). In fact, survey data suggests organizations spend approximately $812 to $1189 annually developing each of their employees (American Society of Training & Development, 2004). This booming industry houses a variety of approaches to growing individuals and teams such as formal training, 360-degree feedback, executive coaching, operational assignments, mentoring programs, guided reflection, action learning, and outdoor challenges (Day & Halpin, 2001).

Of the above approaches to developing teams and their members, formal training programs are by far the most widely utilized (Salas, Burke, & Stagl, 2004). Conceptually, formal training programs, or instructional strategies, consist of theoretically grounded tools and methodologies, which are combined with a set of competencies and training objectives to form an instructional strategy (see Figure 2) (Salas & Cannon-Bowers,

1997). Operationally, training is a planned intervention conducted to enhance the direct determinants of performance, including job relevant knowledge, skills, and volitional choice behaviors (Campbell & Kuncel, 2001). For example, stress exposure training relies upon decades of stress research to develop the personal resources of team members, and thereby a teams' capacity to buffer the effects of stressors.

Best Practice #1: *Blend scientifically grounded tools, methodologies, competencies, and training objectives to create training programs*

A commonality of all training programs is that they include one of three methods for conveying instructional material (i.e., information-based, demonstration-based, practice-based) (Salas & Cannon-Bowers, 2001). Information-based methods deliver facts, concepts, and tips via lectures, handbooks, advanced organizers, and slide-shows. Demonstration-based methods visually depict behaviors, actions, and/or strategies via video or multimedia information technologies. Practice-based methods provide trainees with hands on rehearsal and real time feedback via role-play exercises, moderate fidelity desktop-based computer simulations, or high fidelity virtual reality simulations.

Best Practice #2: *Design training to leverage information-based, demonstration-based, and/or practice-based methods*

Formal training programs target either individuals and/or teams for development. The emphasis in this chapter is on team training; although team members as individuals are often developed first prior to achieving the team level effects team training programs are designed to produce. In fact, it has been suggested that knowledge and skills should be imparted first at the individual level, progress to role dyads, and then extend to the role networks comprising the team (Kozlowski, 1998). In other words, members should first

learn to master their own responsibilities before discovering how their activities mesh with those of their teammates.

Best Practice #3: *Begin by training team members then move on to teams*

While team training programs vary in content (e.g., crew resource management training, stress exposure training, metacognitive training) and method (e.g., information, demonstration, practice), all of these approaches share the similar aim of leveraging principles of learning to create meaningful growth experiences that benefit both individuals and their employer. This speaks to the purposeful nature of training, it is more than a feel good intervention; it is a substantial investment that must be evaluated to make decisions about the continued use of the program and about the trainees exposed to the learning opportunity. In terms of evaluation, training must impart knowledge, affect, and skills and produce results that horizontally and vertically transfer to the workplace (Kozlowski, Brown, Weissbein, Cannon-Bowers, & Salas, 2000). This perspective extends the traditional consideration of the mechanisms that facilitate horizontal transfer from the learning environment to the performance environment, to include a specification of the higher level (e.g., team, organizational) characteristics that enhance the expression of newly developed competencies in context (Kozlowski & Salas, 1997).

Best Practice #4: *Conduct training evaluation to gauge horizontal transfer and vertical transfer*

In regards to evaluating training, organizational stakeholders should also estimate the return on investment from training and utility analysis can be utilized for this purpose (Schmidt, Hunter, & Pearlman, 1982; Boudreau, 1983). Mathieu and Leonard (1987) advanced a time-based utility model which can be applied to estimate the dollar value of

training. This model is noteworthy because it incorporates additional economic parameters that previous formulas have not taken into account such as variable costs, marginal tax rates, and discounting. In contrast to a deterministic utility approach which uses discounted cash flows to arrive at the net present value of an investment, training can also be valued using a real options framework (Nembhard, Nembhard, & Qin, 2005). Modeling the value of training with this latter approach acknowledges the volatility of production dynamics, labor dynamics, task heterogeneity, and workforce heterogeneity.

Best Practice #5: *Estimate the return on investment from training programs*

To this point, we have noted that team training is an effective means of imparting knowledge, skills, and attitudes, and thereby is a useful means of shaping workplace behavior. Increasingly, however, the designers and end-users of team training programs are concerned with developing trainee' knowledge structures such as shared team-interaction mental models (Kraiger, Ford, & Salas, 1993; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). This is in contrast to the traditional emphasis on imparting knowledge of general principles and developing finite skills. Unfortunately, research suggests traditional training programs are often insufficient for fostering new or changing existing mental models (Brigham & Laios, 1974; Morris & Rouse, 1985).

Best Practice #6: *Think twice before utilizing traditional team training programs to impart or change knowledge structures*

Instructional strategies that are designed to impart mental models must go beyond merely presenting task principles in order to emphasize the components that comprise a system, as well as, the interdependencies between those components (Kieras & Bovair, 1984). One instructional strategy that is useful for imparting both team member

knowledge structures and shared team knowledge structures is cross training (Marks, Sabella, Burke, & Zaccaro, 2002). Cross training has been successfully utilized to train both private sector and military teams (see Marks et al., 2002; McCann, Baranski, Thompson, & Pigeau, 2000; Volpe et al., 1996). The results, lessons learned, and best practices that can be distilled from these interventions are discussed in greater detail next.

Best Practice #7: *Use cross training to impart interpositional knowledge, a specific type of shared team-interaction mental model*

A Science of Cross Training

Cross training is a team training strategy that is used to train each of several team members in the tasks, roles, and responsibilities of their teammates (Volpe et al., 1996). By providing team members with insight into, and sometimes practice with, their teammates' duties, cross training serves to develop within team interpositional knowledge, a specific type of shared team-interaction mental model. Mental models are "mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states" (Rouse & Morris, 1986, p.351). These instantiated schemas can be held about a range of areas of team functioning such as a team's task, equipment, team members, and team member interactions (Cannon-Bowers, Salas, & Converse, 1993).

In keeping with the definition advanced by Marks et al. (2002, p.5) in their recent research on cross training, we define a team-interaction mental model as the "content and organization of interrole knowledge held by team members in a performance setting." These cognitive repositories consist of procedural knowledge about how members work together to address problems within a given domain, including information about who

does what and when those actions are most appropriate. Thus, team-interaction models are cognitive states that members draw upon to represent, explain, and predict the content and sequence of interdependent activities during goal accomplishment.

Shared mental model theory posits that members of effective teams hold *complementary*, and sometimes *overlapping* representations of their tasks, equipment, members, and member interactions. Whether the procedural knowledge which comprises shared team-interaction models must be identical (Cannon-Bowers et al., 1993; Klimoski & Mohammed, 1994) or compatible (Kozlowski, Gully, Salas, & Cannon-Bowers, 1996; Wegner, 1995) is an issue beyond the scope of the current paper (see Kozlowski & Klein, 2000). For now, we simply assert that team members who understand the demands faced by their colleagues' will be better positioned to coordinate with their teammates via the provision of assistance or leadership during routine or adaptive performance.

As evidence mounts for the assertion that effective teams have members with shared knowledge (e.g., Entin & Serfaty, 1994; Marks, Zaccaro, & Mathieu, 2000), shared mental models are increasingly viewed as a powerful theoretical explanation for effective team performance. For example, Hemphill and Rush's (1952) research with B-29 aircrews found a detailed understanding of team member positions (i.e., IPK) was correlated with both individual and crew effectiveness. Specifically, IPK was moderately correlated with crew coordination, crew leadership, and crew initiative. Moreover, in a subset of B-29 crews who completed an extended measure of IPK, interpositional knowledge was found to be correlated with a host of performance indices as rated by supervisors. In this sub-sample, IPK was positively correlated with competence on arrival in the unit, effectiveness in working with others, performance under stress, conformity

with standard operating procedures, attitude and motivation, and overall effectiveness.

With respect to unique theoretical developments in cross-training, Fiore et al. (2005) argued that presenting performance information during after-action review via differing *perspectives* represents an important way to potentially build IPK. They devised a technique based upon narrative theory (e.g., Bal, 1997) that would support the development of IPK through a unique form of positional modeling. This form of training may augment views of learning via positional rotation which has been argued to support "greater development of interpositional knowledge, which helps to enhance team functioning in difficult environments that require implicit coordination strategies" (Hollenbeck, DeRue, & Guzzo, 2004, p. 360).

An additional theoretical development has arisen out of the study of groups. This is the notion of transactive memory systems within groups, a concept which may similarly contribute to an improved understanding of cross-training. Team familiarity in the form of "transactive memory systems" is related to successful task execution (e.g., Liang, Moreland, & Argote, 1995) in that these systems are described as long-term memory stores. These are essentially a form of shared mental model whereby a team does not store all aspects of knowledge relevant to the team or to their task, instead members retain who is aware of particular bits of information (Moreland & Myaskovsky, 2000). Fiore and colleagues have argued that this form of shared mental model directly relies on shared episodic memories, that is, they pertain to "interactions with teammates and the scenarios or situations engaged by the team" (Fiore, Salas, Cuevas, & Bowers, 2003, p. 353). Thus, as teams interact, communication and coordination proceeds often with the development of a shared or "common language for describing tasks" (Liang et al., 1995

p. 386). Cross-training techniques relying on understanding and developing the forms of communication patterns that are indicative of the development of shared mental models and to effective team coordination represent important developments in team training.

To this point, our discussion has treated cross training as a unified team training approach when, in fact, it consists of a suite of three interrelated training techniques (i.e., positional clarification, positional modeling, positional rotation). These three techniques differ in regards to the method used to impart knowledge about one's teammates' roles and duties (Blickensderfer, Cannon-Bowers, & Salas, 1998). For example, *positional clarification* is used to transmit a general understanding to a team member about his or her teammate's position via lecture or discussion. *Positional modeling* extends the technique of positional clarification by adding an observational component to the discussion of one's teammates roles and responsibilities. *Positional rotation*, is the most complex form of this instructional strategy, as it combines presentation and modeling with actual practice handling one's teammates' responsibilities and tasks.

The three techniques which comprise cross training can be conceptualized as varying along a continuum of learning outcomes (Marks et al., 2002). At one end of the spectrum is positional clarification, which is used to raise awareness of the type and timing of activities that occur within a team. In contrast, positional rotation is at the other extreme because it is used to build knowledge and skill redundancy. While each of the three types of cross training is used with the common objective of increasing IPK, coordination, and thereby effective performance, each intervention varies in intensity, instructional method, monetary costs, time commitment, practicality, and potential negative transfer of training (Marks et al., 2002).

Best Practice #8: *Consider the pros and cons associated with each of the three cross training techniques in light of growth objectives and contextual constraints*

Because the use of each type of cross training is associated with specific up and downside factors, it is important to cultivate a clear understanding of the effects of these three interventions as separate techniques. This caveat is especially pertinent in light of meta-analytic research findings which provided meager support for the contribution of cross training to the overall effectiveness of team training when it was conceptualized as a monolithic intervention (Salas, Mullen, Nichols, & Driskell, 2002). It is only through illuminating the unique benefits and costs of each of these three techniques that organizations can leverage that understanding to tailor the use of a given type of cross training to meet the specific training objectives in a particular context.

The following subsections begin to delineate the proximal and distal effects of cross training by discussing the findings of prior research initiatives in light of an advanced heuristic framework (see Figure 3). The discussion which follows of the empirical research studies investigating cross training is structured around the three techniques which comprise cross training (i.e., positional clarification, positional modeling, positional rotation). Best practices are extracted from these initiatives and advanced within each subsection. These best practices can be used by organizations conducting cross training to benchmark their current interventions or create new ones.

Positional Clarification

Positional clarification relies upon written or verbal statements presented via lecture or discussion to convey IPK to team members about their teammates' positions. The instructor-trainee interactions which occur during a positional clarification

intervention are conducted to raise the trainee's awareness about the content and timing of the taskwork and teamwork processes his / her teammates enact.

To date, two research studies have investigated the positional clarification approach to cross training (see Cooke, Kiekel, Salas, & Stout, 2003; Marks et al., 2002). The first of these two initiatives examined the effects of positional clarification on shared cognition in two laboratory settings using two different computer-based military simulations (Marks et al., 2002). This line of research also gauged the relationships between shared team-interaction mental models and both performance processes and performance outcomes. The results of the first of the two experiments conducted by Marks and colleagues' (2002) suggested that teams exposed to positional clarification cross training developed a higher percentage of shared team-interaction knowledge than control group teams. The results of this first experiment also suggested that mental model similarity was correlated with backup behavior quantity, backup behavior quality, and team performance outcomes. Moreover, both backup behavior quantity and backup behavior quality were strongly correlated with team performance outcomes. The results of Marks et al.'s (2002) second experiment supported no such linkage between the positional clarification form of cross training and shared team-interaction mental models. The variable effects of positional clarification cross training on shared cognition as found in these two experiments is discussed in greater detail later in this subsection.

Best Practice #9: *Use positional clarification cross training to impart shared team-interaction mental models*

The second initiative undertaken to examine positional clarification was also conducted in a laboratory setting and similarly used a computer-based military simulation

(Cooke et al., 2003). This study compared the usefulness of a conceptual version of cross training to a full version of positional clarification. While the results of this endeavor provided little support for conceptual cross training, they did suggest that positional clarification was an effective means of imparting both taskwork IPK and teamwork IPK. Although positional clarification did not directly affect team performance outcomes, taskwork IPK was predictive of performance outcomes in terms of mission completion rate. This latter finding is similar to the results of Marks et al. (2002), who found that the positional clarification affected proximal phenomena such as shared cognition rather than directly impacting more distal team performance outcomes.

Best Practice #10: *Conceptualize and measure the proximal effects of positional clarification cross training to gauge its impact*

As noted, positional clarification is the least intensive of the three cross training techniques. While positional clarification requires the least commitment of organizational resources in terms of time and money, some evidence suggests that it can be a useful means of imparting shared team-interaction models and thereby performance processes and performance outcomes (Marks et al., 2002). In fact, in the first experiment of a two experiment research study, Marks and colleagues found that positional clarification was as effective as positional modeling in creating shared mental models. Given the additional resources required to conduct positional modeling, these results could be interpreted to suggest that positional clarification is a superior technique. Unfortunately, however, the picture is not so clear. This is because the results of the second experiment of this dual experiment research study suggested positional modeling was more effective at engendering shared team-interaction models than was positional clarification.

One explanation advanced by Marks et al. (2002) for the superiority of positional clarification over positional modeling in their first experiment, and its inferiority in their second experiment, is team member collocation or proximity. In their first experiment, adhoc team members navigated a single helicopter such that they were always collocated. Marks and colleagues' reasoned that positional clarification was sufficient for creating interrole knowledge in this context because members worked closely together in real time on the same tasks. In contrast, members in the second experiment operated one of several tanks which coordinated at a distance to accomplish a mission and thus a more intensive form of cross training was required to create shared cognition. It seems, that as the coordination demands within a team increase (e.g., when a team is distributed) so to does the degree of cross-training required (Travillian, Volpe, Cannon-Bowers, & Salas, 1993).

Best Practice #11: *Take into account the additional ambiguities inherent to distributed performance arrangements when choosing between positional clarification and more intensive forms of cross training*

Positional Modeling

The second form of cross training is positional modeling. The defining characteristics of positional modeling as a cross training technique are demonstration and observation. More specifically, positional modeling includes the presentation of instructional content via an information technology-based medium (i.e., video, DVD) or by the efforts of a training instructor whom demonstrates actions or strategies. Both of these mediums allow trainees to observe either desirable or undesirable behaviors.

To date, two research studies have been conducted to examine the effects of positional modeling on team cognition, processes, and performance outcomes (see

Duncan, Rouse, Johnston, Cannon-Bowers, Salas, & Burns, 1996; Marks et al., 2002). The aforementioned Marks et al. (2002) research study consisted of two experiments, the first of which contrasted positional modeling with positional clarification and a control condition. The results of this first experiment suggested that positional modeling was an effective technique for fostering shared team-interaction mental models. Moreover, mental model similarity was correlated with backup processes and team performance outcomes. The results from Marks and colleagues' second experiment lent additional support to the effects of positional modeling on shared cognition. Specifically, the results of their second experiment suggested positional modeling was a useful technique for developing shared team-interaction knowledge and, in turn, this knowledge was correlated with coordination processes and team performance outcomes.

Best Practice #12: *Use positional modeling cross training to impart to impart shared team-interaction mental models*

Duncan and colleagues' (1996) also examined the effects of positional modeling. This research used Team Model Training (TMT), a one hour computer-based cross training program, to examine the effects of cross training on proximal shared team-interaction mental models and more distal teamwork processes (i.e., backup behavior, leadership, situation assessment). The results of this initiative suggested that teams exposed to TMT developed richer shared mental models and engaged in less communication in terms of asking questions. Those teams exposed to TMT were also more effective at engaging in performance processes, such as backup behavior, leadership, and situation assessment, than control group teams.

Best Practice #13: *Use positional modeling cross training to impart performance processes such as leadership, backup behavior, and situation assessment*

Positional Rotation

Positional rotation is the third type of cross training. This technique combines and extends the main elements of positional clarification (i.e., information presentation) and positional modeling (i.e., information demonstration and observation) by including actual practice in the tasks team members are charged with completing. A treatment is provided in this subsection of the five initiatives which have investigated positional rotation.

One early empirical examination of cross training was conducted with military teams which performed tasks in a simulated combat information center environment (Travillian, Volpe, Cannon-Bowers, & Salas, 1993). Travillian and colleagues used positional rotation to develop the IPK of 120 Naval recruits randomly assigned to 40 three-person teams. The results of this study indicated that cross trained teams were more effective at engaging in teamwork (i.e., coordination, leadership, orientation) and in generating desired team performance outcomes (i.e., accuracy of decisions) than control group teams who only received training on their own specific responsibilities.

Best Practice #15: *Use positional rotation cross training to promote emergent cognitive states such as IPK and team performance processes such as leadership*

Interestingly, the results of this initiative indicated the use of positional rotation was not related to mutual performance monitoring, backup behavior, providing feedback, and communication. The researchers interpreted these findings as suggesting that teams with less IPK spent more time monitoring, communicating, and backing-up their

teammates in order to understand the team's interactional requirements and thereby ascertain the 'big picture' (Travillian et al., 1993).

The above results suggest that cross training serves to increase the frequency of some teamwork processes (i.e., coordination) while concurrently suppressing the use of other processes (i.e., communication). These finding have important implications for training evaluation as well as the appraisal of team performance in the workplace. Performance measurement systems used for training evaluation, team development, and/or compensation decisions must be tweaked to acknowledge that positional rotation leads to increases in some performance processes, while diminishing the use of other processes. For example, teams assessed in a maximum performance context after being exposed to cross training may display less communication. This does not necessarily suggest the team is performing poorly, as members may have drawn from their shared knowledge to switch to an implicit coordination strategy that minimizes the amount of overt communication required to complete a set of tasks (Entin & Serfaty, 1994). Moreover, as is suggested by the results of studies discussed later in this text, research suggests that some facets of communication (i.e., volunteers information) are enhanced by cross training while other facets are unaffected (i.e., requests information) (Cannon-Bowers, Salas, Blickensderfer, & Bowers, 1998; Volpe, Cannon-Bowers, Salas, & Spector, 1996). Thus, the variable effects of cross training on various criterion constructs and facets of criteria are also important considerations for organizational stakeholders.

Best Practice #16: *Revise performance measurement systems to account for the nuanced effects of cross training on team inputs and processes*

A second use of positional rotation is found in research investigating the effects of cross training on the processes and performance of crew dyads engaging in a F-16 aircraft simulation (Volpe et al., 1996). Prior to the performance simulation, crew members received: (1) a 10 minute tape recorded message explaining how the simulation equipment operated, (2) written instructions about the task they were about to perform, and (3) a description of the roles and responsibilities of their teammates. This kind of information is identical to the instructional content that could be provided in positional clarification cross training. In addition to this information, each activity the crew members were expected to perform (e.g., aircraft maneuvering, radar locking, weapon selection, weapon firing) was demonstrated by the instructor using the flight simulation to depict appropriate actions. This type of demonstration is akin to positional modeling. Finally, each crew member practiced their own tasks as well as the tasks assigned to their teammate (i.e., positional rotation).

The results of Volpe et al.'s (1996) research suggested that cross trained crews exhibited more effective teamwork processes and produced more effective performance outcomes than their control group counterparts. For example, those dyads that received positional rotation cross training were more effective at communicating in terms of volunteering more information without being asked. Crews exposed to positional rotation were also more effective at maneuvering enemy targets into range, locking on to those targets, and destroying them in an expeditious manner. Cross trained crews also received higher instructor ratings on team competency and overall team quality.

Best Practice #17: *Use positional rotation cross training to reduce the latency with which targeted team performance outcomes are realized*

In a replication and extension of Volpe et al.'s (1996) research, Cannon-Bowers et al. (1998) examined the effects of positional rotation cross training on team processes and performance outcomes under varying workload conditions (i.e., low and high). Using the Tactical Naval Decision Making System to simulate Combat Information Center operations, positional rotation cross training enhanced both IPK and the amount of information volunteered by teams. Cannon-Bowers and colleagues' results also supported their expectations about the interaction of positional rotation and workload. Specifically, those teams that received positional rotation were able to leverage that advantage during the high workload condition to enact higher quality processes, engage more targets correctly, and do so in less time, than teams who did not receive cross training. In contrast, there was no difference between cross trained and control group teams in the low workload condition in terms of process quality or performance outcomes.

Best Practice #18: *Consider the workload typically experienced by a team, as positional rotation cross training is most effective when tasks are demanding*

The effects of a variation of positional rotation cross training, labeled *experiential cross training in a team context*, on team performance have also been investigated using a simulated Naval surveillance task (McCann et al., 2000). McCann and colleagues' sought to examine whether training team members while performing the task as a team was a viable training technique. This approach is in contrast to prior research which used a cross training intervention to train individual team members and only later captured the performance of intact teams comprised of trained members. In addition, teams were reconfigured (i.e., each member was shifted to a new position) during the experiment to

mimic the turnover and absenteeism of personnel which teams in the workplace often experience as part of their ongoing operations.

Surprisingly, McCann et al.'s (2000) research results suggest that the only difference between cross trained and untrained teams in terms of team processes was the fact that the former made slower decisions than the later. Interestingly, however, McCann and colleagues' results suggested that teams exposed to experiential cross training were more effective at coping with personnel reconfigurations than teams who were not exposed to positional rotation. In fact, teams in the experiential cross training condition experienced no performance decrements as a result of the shift in personnel.

Best Practice #19: *Train intact teams via experiential positional rotation cross training to minimize the performance decrements that result from turnover*

The most recent examination of the effects of positional rotation cross training was conducted by Marks et al. (2002). These researchers examined the effects of positional rotation on shared cognition and the relationships between shared cognition and team processes and performance outcomes. The results of this experiment suggested that teams exposed to positional rotation cross training developed mental models with a higher percentage of shared team-interaction knowledge than those teams exposed to positional clarification. Moreover, team-interaction mental models were positively correlated with team coordination and performance outcomes in terms of the number of enemy installations captured and rebuilt on the battlefield.

Best Practice #20: *Use positional rotation cross training to foster team processes such as coordination and thereby valued team performance outcomes*

Executive Summary

There are many approaches to developing human capital which practitioners can utilize to develop teams and their members. This chapter reviewed the conceptual basis and supporting empirical evidence for one of these techniques, cross training. Cross training can be leveraged to develop shared team-interaction mental models; or emergent cognitive states which provide a structure for collecting and organizing information that is subsequently utilized to anticipate and execute the core processes of team performance. As performance unfolds and compiles over time, valued outcomes are achieved and stakeholders, both internal and external to a team, form effectiveness judgments about a team's success and its need for further exposure to developmental opportunities.

The totality of empirical research findings reviewed herein strongly support the use of cross training to develop shared team-interaction mental models. In fact, cross training is a useful means of imparting both teamwork and taskwork models (see Figure 3). The evidence reviewed suggested the quantity and quality of these shared cognitive reservoirs were positively correlated with a wide range of performance processes such as coordination, communication (volunteering information but not requesting information), backup behavior, leadership, and situation assessment. In turn, these episodic processes were positively correlated with a myriad of performance outcomes such as the number of missions completed, decision accuracy, and process quality (see Figure 3). Moreover, the use of cross training reduced decision latency and information access time.

Given the findings reviewed in this chapter, it seems cross training is a valuable approach for turning a team of experts into an expert team. It is important to remember, however, that cross training is a suite of interrelated techniques rather than a single

intervention; and each of these approaches is uniquely grounded in the science of team training. In an effort to help practitioners maximize their limited resources, this chapter advanced 20 best practices which can be leveraged by stakeholders when instituting positional clarification, positional modeling, and positional rotation programs in their organizations. Diligently following the advanced practices will help ensure the integrity, viability, and ultimately the utility of capital allocations for creating expert teams.

Acknowledgement

The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This work was supported in part by funding from the Army Research Laboratory's Advanced Decision Architecture Collaborative Technology Alliance (Cooperative Agreement DAAD19-01-2-0009). This work was also supported in part by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621.

References

American Society of Training and Development (ASTD). (2003). "STD 2003 State of the Industry Report Executive Summary", Retrieved June 3, 2004, from http://www.astd.org/NR/rdonlyres/6EBE2E82-1D29-48A7-8A3A-357649BB6DB6/0/SOIR_2003_Executive_Summary.pdf.

Ancona, D. & Chong, C. L. (1999). Cycles and synchrony: The temporal role of context in team behavior. *Research on Managing Groups and Teams*, 2, 33-48.

Baker, C. V. (1991). *The effects of inter-positional uncertainty and workload on team coordination skills and task performance*. Unpublished doctoral dissertation, University of South Florida, Tampa, FL.

Bal, M. (1997). *Narratology: Introduction to the theory of narrative.* Toronto: University of Toronto Press.

Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (1998). Cross-training and team performance. In J. A. Cannon-Bowers, & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 299-311). Washington, DC: APA.

Boudreau, J. W. (1983). Economic considerations in estimating the utility of human resource productivity improvement programs. *Personnel Psychology, 36*, 551-576.

Brigham, F. & Laios, L. (1974). Operator performance in the control of a laboratory processing plant. *Ergonomics, 18*, 53-66.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (in press). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology.*

Campbell, J. P. & Kuncel, N. R. (2002). Individual and team training. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (pp. 272-312). London, England: Sage.

Cannon-Bowers, J. A. & Salas, E. (1990, April). Cognitive psychology and team training: Shared mental models in complex systems. In K. Kraiger (Chair), *Cognitive representations of work.* Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, Miami, FL.

Cannon-Bowers, Salas, E., & Converse, S. A. (1991). Cognitive psychology and team training: Shared mental models of complex systems. *Human Factors Society Bulletin, 37*, 1-4.

Cannon-Bowers, Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan Jr. (Ed.), Current *issues in individual and group decision making* (pp. 221-246). Hillsdale, NJ: Erlbaum.

Cannon-Bowers, J. A., Salas, E., Blickensderfer, E., & Bowers, C. A. (1998). The impact of cross-training and workload on team functioning: A replication and extension of initial findings. *Human Factors, 40*, 92-101.

Carnevale, A. P., Gainer, L. J., & Villet, J. (1990). *Training in America*. San Francisco: Jossey-Bass.

Cooke, N. J., Cannon-Bowers, J. A., Kiekel, P. A., Rivera, K., Stout, R. J., & Salas, E. (2000). Improving teams' interpositional knowledge through cross-training. *Proceedings of the IEA 2000/HFES 2000 Congress* (pp. 390-393). San Diego.

Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C., & Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research, and Practice, 7*, 179-199.

Day, D. V. & Halpin, S. M. (2001). *Leadership development: A review of industry best practices* (Technical Report Number 1111). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Duncan, P. C., Rouse, W. B., Johnston, J. H., Cannon-Bowers, J. A., Salas, E., & Burns, J. J. (1996). Training teams working in complex systems: A mental model-based approach. In W. B. Rouse (Ed.), *Human/technology interaction in complex systems* (Vol. 8, pp. 173-231). Greenwich, CT: JAI Press.

Fiore, S. M., Johnston, J. H., & McDaniel, R. (2005). *Applying the narrative form and XML metadata to debriefing simulation-based exercises*. Proceedings of the Human Factors and Ergonomics Society 49th annual meeting (pp. 2135-2139). Orlando, FL.

Hemphill, J. K. & Rush, C. H. (1952). *Studies in aircrew composition: Measurement of cross-training in B-29 aircrews* (AD B958347). Columbus: Ohio State University, Columbus Personnel Research Board.

Hollenbeck, J.R., DeRue, D.S., and Guzzo, R. (2004). Bridging the gap between I/O research and HR practice: Improving team composition, team training and team task design. *Human Resource Management, 43, 4,* 353-366.

Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. A. (1964). *Organizational stress: Studies in role conflict and ambiguity*. New York: Wiley.

Kieras, D. E. & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Psychology, 8*, 255-273.

Klein, G. A. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in man-machine systems research* (Vol. 5, pp. 47-92). Greenwich, CT: JAI Press.

Klimoski, R., & Mohammad, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20*, 403-437.

Kozlowski, S. W. J. (1998). Training and developing adaptive teams: Theory, principles, and research. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (91-114). DC: APA.

Kozlowski, S. W. J., Brown, K., Weissbein, D., Cannon-Bowers, J., & Salas, E. (2000). A multilevel approach to training effectiveness: Enhancing horizontal and vertical transfer. In K. Klein, & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organization* (pp. 157-210). San Francisco, CA: Jossey-Bass.

Kozlowski, S. W. J., Gully, S. M., Salas, E., & Cannon-Bowers, J. A. (1996). Team leadership and development: Theory, principles, and guidelines for training leaders and teams. In M. Beyerlein, S. Beyerlein, & D. Johnson (Eds.), *Advances in interdisciplinary studies of work teams: Team leadership* (Vol. 3, pp. 253-292). Greenwich, CT: JAI.

Kozlowski, S. W. J. & Klein, K. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3-90). San Francisco: Jossey-Bass, Inc.

Kozlowski, S. W. J., & Salas, E. (1997). A multilevel organizational systems approach for the implementation and transfer of training. In. J. K. Ford (Ed.). *Improving training effectiveness in work organizations* (pp. 247-287). Mahwah, NJ: Lawrence Erlbaum Associates.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.

Liang, D., Moreland, R., & Argote, L. (1995). Group versus individual training and group performance: The mediating factor of transactive memory. *Personality & Social Psychology Bulletin, 21*, 384-393.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team process. *Academy of Management Review, 26*, 356-376.

Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology, 87*, 3-13.

Marks, M. A., Zaccaro, S. J., & Mathieu, J. E. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *Journal of Applied Psychology, 85*, 971-986.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*, 273-283.

Mathieu, J. E. & Leonard, R. L. (1987). Applying utility concepts to a training program in supervisory skills: A time-based approach. *Academy of Management Journal, 30*, 316-335.

Mathieu, J. E. & Schulze, W. (in press). The influence of team knowledge and formal plans on episodic team process → performance relationships. *Academy of Management Journal*.

McCann, C., Baranski, J. V., Thompson, M. M., & Pigeau, R. A. (2000). On the utility of experiential cross-training for team decision making under time stress. *Ergonomics, 43*, 1095-1110.

McIntyre, R. M. & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 149-203). San Francisco, CA: Jossey-Bass.

Moreland, R. L. & Myaskovsky, L. (2000). Exploring the performance benefits of group training: Transactive memory or improved communication? *Organizational Behavior & Human Decision Processes, 82*, 117-133.

Morris, N. M. & Rouse, W. B. (1985). The effects of type of knowledge upon human problem solving in a process control task. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-15*, 698-707.

Nembhard, D. A., Nembhard, H. B., & Qin, R. (2005). A real options model for workforce cross-training. *The Engineering Economist, 50*, 95-116.

Orasanu, J., & Salas, E. (1993). Team decision making in complex environment. In G. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 327-345). Norwood, NJ: Ablex.

Saavedra, R., Earley, P. C., & Van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology, 1,* 61-72.

Salas, E., Burke, C. S., & Stagl, K. C. (2004). Developing teams and team leaders: Strategies and principles. In D. Day, S. J. Zaccaro, & S. M. Halpin (Eds.), *Leader development for transforming organizations*. Mahwah, NJ: Erlbaum & Associates, Inc.

Salas, E. & Cannon-Bowers, J. A. (1997). Methods, tools, and strategies for team training. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 249-280). Washington, DC: APA.

Salas, E. & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471-499.

Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (1997). How can you turn a team of experts into an expert team?: Emerging training strategies. In C. E. Zsambok, & G. Klein (Eds.), *Naturalistic decision making* (pp. 359-370). Mahwah, NJ: Lawrence Erlbaum Associates.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward and understanding of team performance and team training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance*. Norwood, NJ: Ablex Publishing.

Salas, E., Mullen, B., Nichols, D. R., & Driskell, J. E. (2002, April). What contributes to the effective design of team training? In A. Towler (Chair), *Current issues in training design and evaluation*. Symposium presented at the 17th annual conference of the Society for Industrial and Organizational Psychology, Toronto.

Salas, E., Priest, H. A., Stagl, K. C., Sims, D. E., & Burke, C. S. (2006). Work teams in organizations: A historical reflection and lessons learned. In L. Koppes (Ed.), *Historical perspectives in industrial/organizational psychology* (chapter 17). Mahwah, NJ: Lawrence Erlbaum Associates.

Salas, E., Stagl, K. C., Burke, C. S. & Goodwin, G. F. (2006). Fostering team effectiveness in organizations: Toward an integrative theoretical framework of team performance. In J. W. Shuart, W. Spaulding, & J. Poland, (Eds.), *Modeling complex systems: Nebraska Symposium on Motivation, 52,* Lincoln, NE: University of Nebraska Press.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1982). Assessing the economic impact of personnel programs on work-force productivity. *Personnel Psychology, 35*, 333-347.

Stagl, K. C., Salas, E., & Burke, C. S. (in press). Best practices in team leadership: What team leaders do to facilitate team effectiveness. In J. A. Conger & R. E. Riggio (Eds.), *The practice of leadership* (pp.). John Wiley & Sons.

Stagl, K. C., Salas, E., Rosen, M. A., Priest, H. A., Burke, C. S., Goodwin, G. F., & Johnston, J. H. (in press). Distributed team performance: A multilevel review of distribution, demography, and decision-making. In F. Yammarino & F. Dansereau (Eds.), *Multi-level issues in organizations*, 5. X, X: Elsevier.

Sundstrom, E., de Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist, 45,* 120-133.

Sundstrom, E., McIntyre, M., Halfhill, T., & Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics, 4*, 44-67.

Tannenbaum, S. I., Beard, R. L., & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. In K. Kelley (Ed.), *Issue, theory, and research in industrial/organizational psychology* (pp. 117-153). Amsterdam: Elsevier.

Travillian, K. K., Volpe, C. E., Cannon-Bowers, J. A., & Salas, E. (1993). *Cross-training highly interdependent teams: Effects on team process and team performance*. Proceedings of the 37th annual Human Factors and Ergonomics Society conference (pp. 1243-1247). Seattle, Washington.

Volpe, C., Cannon-Bowers, J. A., Salas, E., & Spector, P. E. (1996). The impact of cross-training on team functioning: An empirical investigation. *Human Factors, 38*, 87-100.

Wenger, D. M. (1995). A computer network model of transactive memory. *Social Cognition, 13*, 319-339.

Figure 1.
Team Effectiveness Framework
Adapted from (Tannenbaum, Beard, & Salas, 1992)

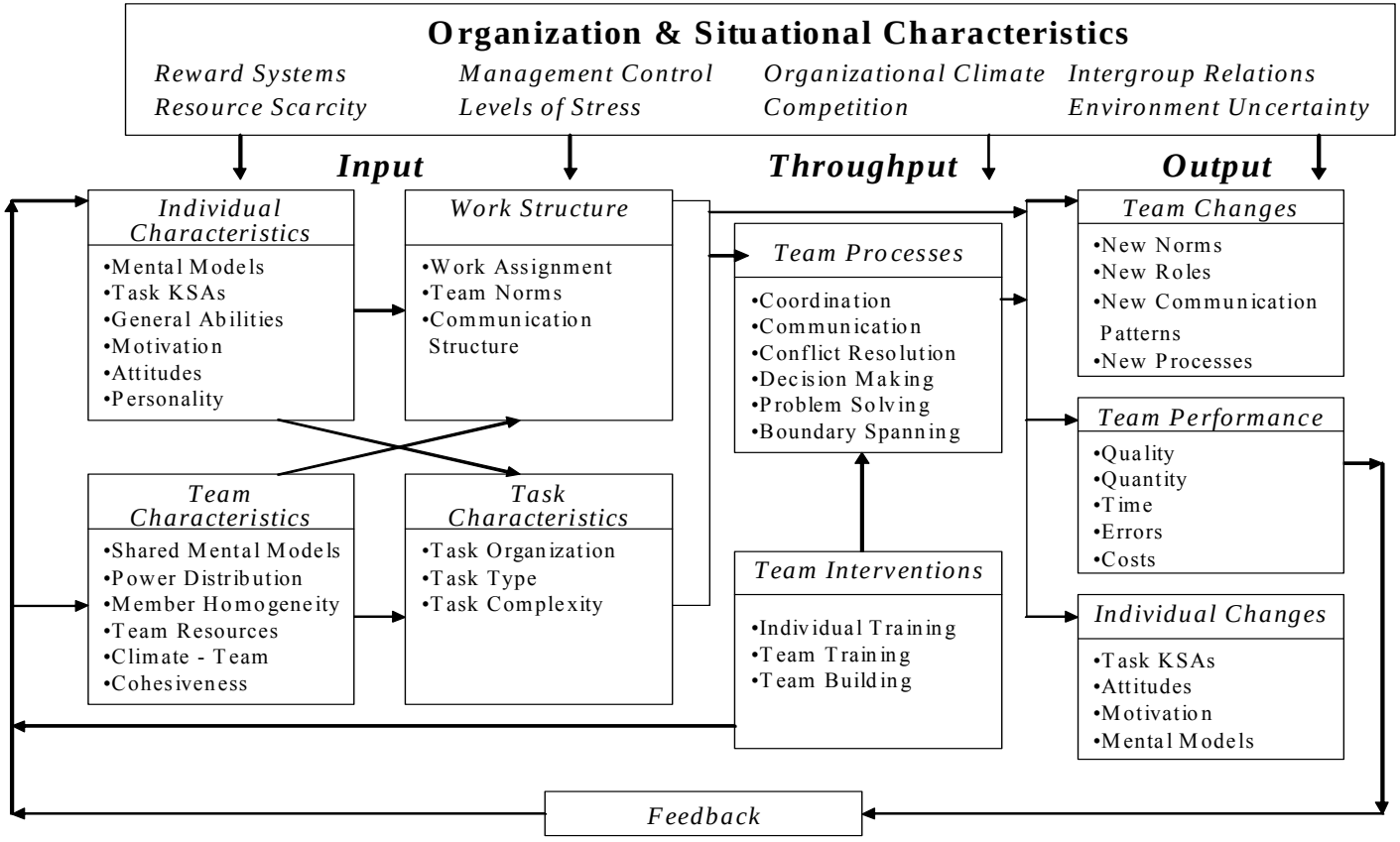


Figure 2.
Model of Structure of Team Training
Adapted from (Salas & Cannon-Bowers, 1997)

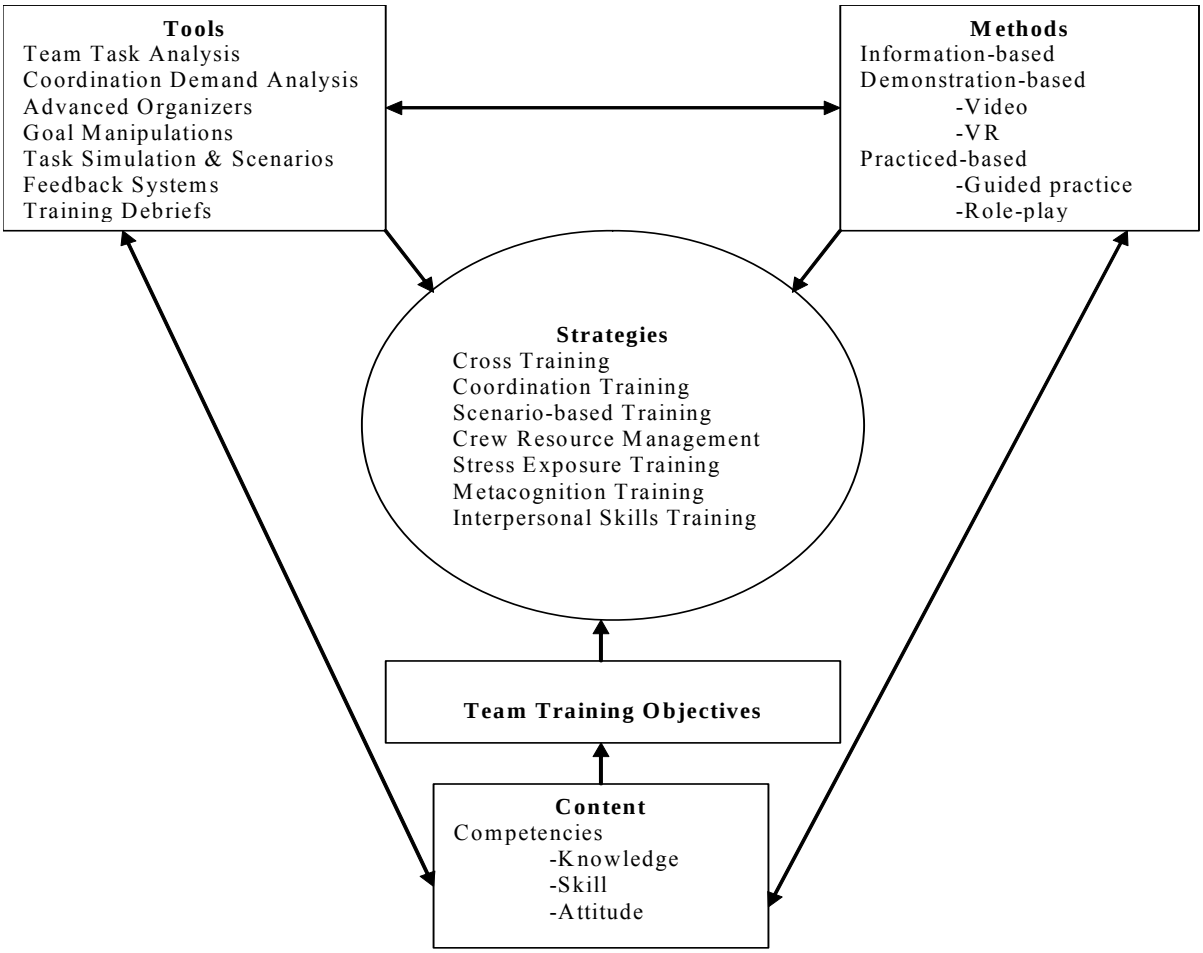
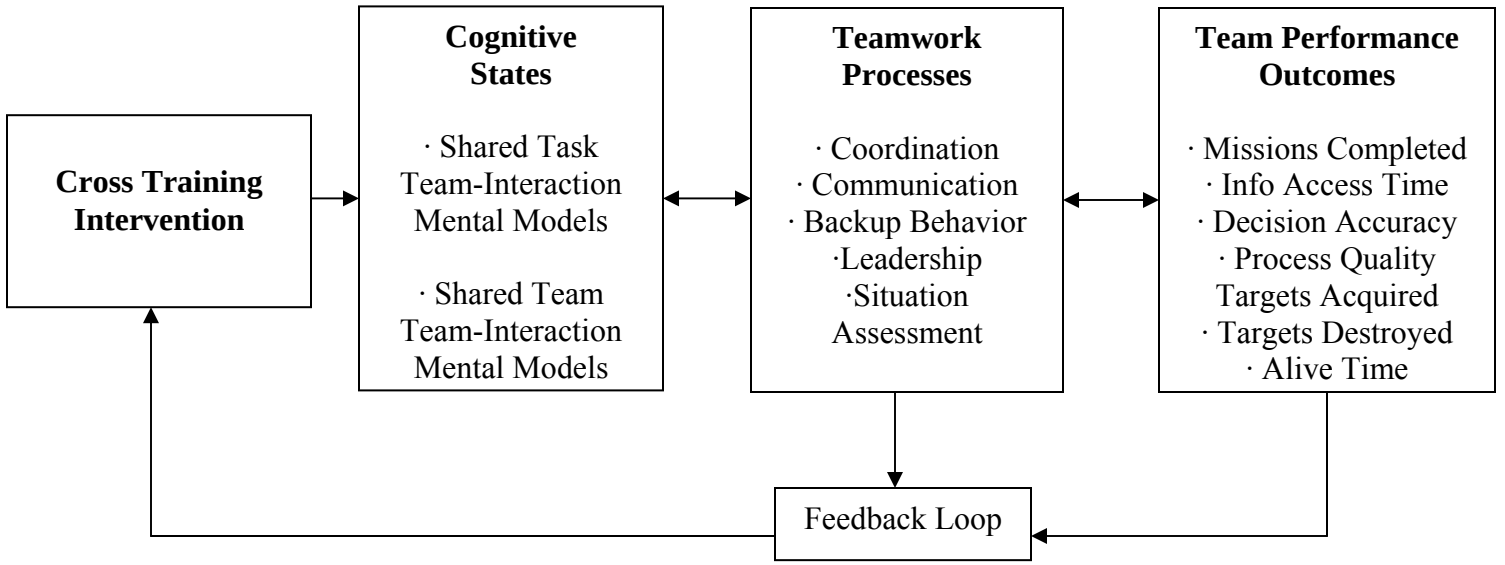


Figure 3.
Heuristic Framework of Cross Training Effects



Professional Biographies

Stephen M. Fiore is faculty with University of Central Florida's Cognitive Sciences Program and Director of the Consortium for Research in Adaptive Distributed Learning Environments at UCF's Institute for Simulation and Training and Team Performance Laboratory. He earned his Ph.D. degree (2000) in Cognitive Psychology from the University of Pittsburgh, Learning Research and Development Center. He maintains a multidisciplinary research interest that incorporates aspects of cognitive, social, and organizational psychology in the investigation of learning and performance in individuals and teams. Dr. Fiore has published in the area of learning, memory, and problem solving at the individual and the group level. He has helped to secure and manage over $5 Million in research funding from organizations such as the National Science Foundation, the Transportation Security Administration, the Office of Naval Research, and the Air Force Office of Scientific Research.

Eduardo Salas is Trustee Chair and Professor of Psychology in the Department of Psychology and the Institute for Simulation and Training at the University of Central Florida. He has over 20 years of research experience on team training, simulation-based training, and team effectiveness and has published over 300 articles, chapters, and papers on those topics.

Kevin C. Stagl is a Research Scientist at the University of Central Florida's Institute for Simulation and Training (IST). At IST, his research addresses the spectrum of team issues with an emphasis on leadership, performance, development, and adaptation. Prior to joining IST, Kevin spent 5 years as a member of an organizational consultancy where he provided clientele with human capital decision support.

Understanding Performance, Workload, and Stress: An Individual Differences Perspective

J.L. Szalma and P.A. Hancock

The modern emphasis on human-centered-design leads to questions regarding the characteristics of the human operator and how these influence and, ultimately, are influenced by, characteristics of the environment; particularly the task itself. However, research on human performance has yet to fully exploit the potential of an individual differences perspective to improve both theoretical understanding of performance under stress as well as practical implications for interface design. In this paper we summarize a series of empirical studies on individual differences in human performance, workload, and stress response in cognitive and perceptual tasks. It is suggested that incorporation of an individual differences perspective can improve two influential theories of stress and performance, the compensatory control model (Hockey, 1997) and the maximal adaptability model (Hancock & Warm, 1989). Although these theories recognize the importance of individual differences, they do not provide a framework describing the links between individual difference variables and the components of their respective models, nor do they indicate the mechanism by which individual difference variables exert their effects. Incorporation of individual differences into these models could greatly enhance our understanding of stress and performance by elucidating how variation across and within individuals influences human performance in stressful environments.

Fuzzy Signal Detection Theory and Human Performance: A Review of Empirical Evidence for Model Validity

J.L. Szalma and P.A. Hancock
University of Central Florida

Operators of complex systems must detect critical signals quickly and accurately under conditions of high workload and stress. Traditional Signal Detection Theory (SDT) is a powerful tool for quantifying system performance, but it is not without limitations. Specifically, in the SDT model the state of the world is forced into crisp mutually exclusive categories (i.e., signal versus non-signal), which may not be accurate representations of the true states of the real world. Fuzzy Signal Detection Theory (FSDT) combines traditional SDT with Fuzzy Set Theory to extend signal detection analysis beyond the traditional crisp, categorical model. FSDT allows for events to simultaneously be in more than one state (e.g. signal and non-signal), so that a particular instantiation of a stimulus has properties of both 'signal' and 'non-signal.' Consequently, FSDT can be employed in applied settings where the degree to which an event is a signal for detection may vary and decisions must be made under conditions of high uncertainty. In this paper we describe recent empirical efforts to test the assumptions of the model and its utility for assessing human performance in complex signal detection tasks.

Proposal title: A Meta-Analysis of Periodic Noise Stress on Human Performance

Author(s) and current affiliations: Ross, J.M., Conway, G.E., Szalma, J.L., Saxton, B., Braczyk, A. & Hancock, P.A.

E-mail and mailing address for presenting author: jennifer.ross@peostri.army.mil
Jennifer Ross
4440 Stonefield Drive
Orlando, FL 32826

Proposal focus: empirical (meta-analysis)

Presentation type: lecture

Student or non-student work: Non-student work

Abstract (150 word maximum): Noise stress, commonly defined as unwanted sound, is perhaps the most ubiquitous stressor that human beings experience. In recent years, there have been numerous studies on noise stress, conducted in the behavioral and social sciences, leading this area to be one of the most studied domains in stress and human performance. Unfortunately, the findings in the noise stress literature are often at odds with one another. The following report addresses the first-step at consistently and quantitatively summarizing the noise stress literature by examining human performance under continuous (periodic) noise in an empirical meta-analysis. After summarizing the effect-size statistics from 73 studies, it was demonstrated that there is a decrement in performance for periodic noise stress overall ($g = -0.19$), but that this effect is highly influenced by task category (e.g., cognitive, motor, etc.) and type of dependent measure (i.e., speed or accuracy).

Summary (1500 word-maximum):

INTRODUCTION

The impact of periodic noise stress on human performance appears to be a straightforward question, but in reality it is quite difficult to answer. The literature abounds with a multitude of alternate findings concerning noise's impact on human performance. Whether noise helps, hinders, or has any effect at all, depends on a host of moderating variables that influence this relationship. This diversity of results has sparked endless controversy on the actual impact of noise on performance, fortunately a new statistical method known as meta-analysis has emerged that can quantitatively summarize the state of noise-stress literature and the impact of moderator variables (e.g., task characteristics, noise characteristics, and population characteristics) on human performance.

Meta-analysis is a form of statistical analysis in which a collection of analysis results from individual studies are mathematically combined for the purpose of integrating the findings to obtain a global picture (Glass, 1976). In the following meta-

analysis, we analyzed the effect of periodic noise stress on human performance and the effects of the moderators: task type and dependent measure (i.e., accuracy or speed). For our analysis task types were based upon four sub-divisions: perceptual, cognitive, psychomotor, and communication (See Table 1). These tasks and their descriptive characteristics were based upon a basic information-processing stage differentiation (cf., Lachman, Lachman, & Butterfield, 1979; Wickens & Hollands, 1999). We choose task type and dependent measure for our preliminary analysis because of strong literature support to indicate that these would be important moderating variables in the noise and human performance relationship. While these moderators have been previously examined on a qualitative level in literature reviews by other researchers (e.g., Smith, 1989) these effects have not been examined at this level of detail using the effective quantitative method of meta-analysis.

Table 1. *Task Types Used and Descriptive Examples.*

Task Type	Description	Example
Perceptual	Requires using the eyes to identify or separate targets or objects.	Change blindness study
Cognitive	Requires processing information mentally and reaching a conclusion.	Arithmetic problems
Psychomotor	Requires moving the body to perform an action (may be discrete or continuous).	Tracking tasks
Communication	Requires either reading text or numbers, writing text or numbers, or talking or listening to someone.	Writing the letter p repeatedly

METHOD

In total 423 research reports (journal articles, dissertations, etc.) were obtained by consulting PsycINFO, subject-matter experts (SMEs), and the reference sections of papers obtained via the first two methods. Each report was then queried in regards to four criteria for inclusion in the meta-analysis:

1) Each study had to present empirical findings that dealt with continuous noise stress (i.e., review papers were excluded as were intermittent noise stress studies)
2) A lower dB control condition was required for comparison with a higher dB noise condition.
3) The dependent variable must measure performance of a healthy adult human participant.
4) Each study had to present enough information to obtain effect size statistics

Of the initial reports those satisfying the criteria for inclusion were 73 papers, resulting in an inclusion rate of 17%. It is important to note that rejecting a number of primary studies in a meta-analysis represents a common occurrence and is necessary to ensure meaningful combination of effect-sizes across studies.

The Calculation of Effect Sizes and Variance Estimates

Effect-size, specifically Hedges' *g* in this analysis, is a standardized mean difference between the experimental and control conditions (see Hedges & Olkin, 1985). Effect-size statistics are used in meta-analysis to allow culmination of effects across different tasks and measures. In calculating the effect size, the sign of the effect size was controlled to ensure that a positive score represented improvement in performance in the experimental group relative to the control group, whereas a negative g-score indicated performance impairment.

Analysis of Results

In analyzing the results of a meta-analysis it is accepted practice to first obtain a global result and then to hierarchically analyze at varying levels of depth within the construct of interest. In the current analysis we collapsed over common participant samples at each level of the analysis to avoide violating the assumed independence between effect-size estimates. The culminated effect-size statistics were then computed based on formulas available in Hedges and Olkin (1985). For comparison purposes, it is helpful to keep in mind Cohen's (1988) guidelines to interpreting effect sizes as small ($g \leq 0.20$), medium ($g \approx 0.50$), or large ($g \geq 0.80$).

In addition to the weighted mean effect size, two variance estimates were computed: variability due to sampling error (s_e^2) and variability of the effect sizes (s_g^2). These values were used to estimate the variability due to differences in the population effect sizes, δ (s_δ^2). A large s_δ^2 indicates that there is variability among the observed effect sizes that cannot be accounted for by sampling error, and that there are likely to be one or more variables moderating the magnitude of the effect in question.

RESULTS

The first result is for the global analysis, drawn from 73 total studies, and shows a small effect (g = -0.24), indicating that performance is degraded under periodic noise (See Table 2). The next step of our meta-analysis was to observe the degree to which noise impairs performance according to the moderating factors of task type and dependent measure. To accomplish this, we initially differentiated the effects according to the task type (See Table 3).

Table 2. *Global Analysis of Periodic Noise Stress*

Category	k	Average g	s_g^2	s_e^2	s_d^2	lower limit of 95% CI	upper limit of 95% CI
Global g	73	-0.24	0.35	0.13	0.21	-0.33	-0.16

Table 3. *Moderator Analysis of Task Category*

Category	k	Average g	s_g^2	s_e^2	s_d^2	lower limit of 95% CI	upper limit of 95% CI
Task Type							
Perception	20	-0.09	0.23	0.07	0.16	-0.2	0.02
Cognitive	49	-0.16	0.32	0.12	0.2	-0.26	-0.06
Psychomotor	6	-0.35	0.25	0.29	0	-0.78	0.08
Communication	8	-0.57	0.19	0.33	0	-0.97	-0.17

It is evident from Table 3 that zero is contained within the confidence interval of *perception* and *psychomotor*, indicating that the effect of periodic noise on these two categories appears to be negligible. With respect to the *cognitive* task category, the average effect size is g= -0.16. This effect is considered to be stable because of the large number of qualifying studies (k=49), and therefore represents a small deleterious effect of periodic noise on cognitive performance. As is evident from Table 3, zero is not contained in the confidence interval, so the magnitude of the effect is not due to sampling error. Eight studies were found that examined *communication* performance under noise, with a medium effect (g= -0.57) reflecting degraded performance under noise. Further, this effect is homogeneous, as the sampling error accounts for all the variance in effect sizes (see Hunter & Schmidt, 2004).

In regards to the dependent measure the effects of periodic noise stress was divided between *speed of response* versus *accuracy of response* (see Table 4). In this division, noise had a small negative effect on both speed (g= -0.13) and accuracy (g=-0.19). Both confidence intervals excluded zero indicating a true negative effect of periodic noise across both speed and accuracy measures.

Table 4. *Moderator Analysis of Dependent Measure*

Category	k	Average g	s_g^2	s_e^2	s_d^2	lower limit of 95% CI	upper limit of 95% CI
Dependent Measure							
Speed	22	-0.13	0.24	0.06	0.18	-0.24	-0.03
Accuracy	57	-0.19	0.47	0.19	0.27	-0.31	-0.08

In order to determine if the negative speed and accuracy noise effects were dependent upon task type, the interaction of these effects was also examined (See Table 5). Only three effects were found, they were accuracy of communication and speed and accuracy of cognitive tasks. In regards to cognitive tasks periodic noise was found to degrade performance on performance *accuracy* (a small effect, g= -0.14) and speed of responses (a small effect, g= -0.22). Note that the 95% CI for these effects do not contain zero, indicating a true effect. For communication studies *accuracy*, yielded a large negative effect (g= -1.01). While, the sampling error variance was unusually large, the 95% confidence interval did not contain zero, indicating a true negative effect for periodic noise on communication accuracy. However, this effect must be considered with caution due to the low number of studies (k=6).

Table 5. *Moderator Analysis of Task Category within each Dependent Measure Category*

Category	k	Average g	s_g^2	s_e^2	s_d^2	lower limit of 95% CI	upper limit of 95% CI
Speed							
perception	10	-0.03	0.25	0.05	0.21	-0.17	0.1
cognitive	8	-0.22	0.15	0.07	0.08	-0.4	-0.03
psychomotor*	2	-0.97					
communication**	2	0.03					
Accuracy							
perception	13	-0.12	0.18	0.09	0.09	-0.28	0.04
cognitive	39	-0.14	0.47	0.17	0.3	-0.26	-0.01
psychomotor	5	-0.19	0.51	0.14	0.37	-0.52	0.14
communication	6	-1.01	0.15	1.12	0	-1.85	-0.16

Note. * The two effects sizes are -0.60 (van Gemmert & van Gaelen, 1997) and -0.98 (van Levy-Leboyer & Moser, 1988); ** The two effects sizes are 0.51 (van Gemmert & van Gaelen, 1997) and -0.41 (Davies & Davies, 1975).

DISCUSSION

The overall effect for periodic noise was in the expected direction of a small negative effect (g = -0.24), indicating that performance is indeed degraded under continuous noise. In breaking down noise stress by task type, we found that the two higher-order task types (i.e., cognitive and communication tasks) were those that suffered under the impact of periodic noise, while lower-order tasks (i.e., perception and psychomotor tasks) demonstrated no impact of noise stress. In analysis of speed and accuracy, both measures showed the expected small negative effect (g = -0.13 and g = -0.19 respectively). However, more detailed analysis again supported the finding that continuous noise hampers only higher-order tasks, in that the only true negative effects (those not having zero in the confidence interval) were for cognition and communication for accuracy, and for cognitive in speed. Unfortunately, there is a gap in the literature and not enough studies to suitably estimate the effect of periodic noise on speed of response in communication tasks. This selective impact of noise upon higher-order tasks can be considered in light of Mandler's (1979) work in which he speculated that stress produces noise in the cognitive system which competes with task demands for limited cognitive, that is, conscious resources. Thus, in Mandler's theory, those memory processes that rely on conscious elaboration of input and representations, namely explicit memory processes used here in the higher-order tasks of cognitive and communication tasks, should be especially degraded by stress. Meanwhile, tasks requiring less conscious resources (e.g., perceptual and psychomotor tasks) would be less impacted by cognitive noise.

REFERENCES: Available Upon Request

A META-ANALYSIS OF PERFORMANCE RESPONSE UNDER THERMAL STRESS

P.A. Hancock, J.M. Ross, and J.L. Szalma

Department of Psychology, and
Institute of Simulation and Training

University of Central Florida
Orlando, FL 32816

Correspondence Address: P.A. Hancock
3100 Technology Pkwy, Suite 337
Orlando, Fl 32826-0544

Tel: (407) 823-2310
FAX: (407) 823-5862
Admin: (407) 823-0383
Email: phancock@pegasus.cc.ucf.edu
Web: www.mit.ucf.edu

Key Words: Thermal Stress, Human Performance, Meta-Analysis

ABSTRACT

Objective

The objective for this work was to quantify the effect of thermal stress on human performance.

Background

Most reviews of the effect of environmental stressors on human performance are qualitative. A quantitative review can aid in advancing both theory and practice.

Method

Meta-analytic methods were applied to the available literature on thermal stress and performance. Two-hundred-ninety-one references were collected, and forty-nine publications met selection criteria. These studies provided 528 effect sizes.

Results

Analyses confirmed a substantial negative effect on performance associated with thermal stress. Effect-size for heat was comparable to that for cold temperatures. Cognitive performance was least affected, while both psychomotor and perceptual task performance were degraded to a greater degree. Other moderating variables were identified and observed to mediate the effect of thermal stress.

Conclusion

Results confirmed the importance of task type, exposure duration, and stressor intensity as key variables impacting how temperature affects performance. Results were consistent with a current theory that stress forces the individual to allocate attentional resources to appraise and cope with the threat which reduces the capacity to process task-relevant information. This represents a maladaptive extension of the narrowing strategy which acts to maintain stable levels of response when stress is first encountered.

Application

These quantitative estimates can be used to design tolerance limits for different task types. While results indicate the necessity for further empirical research to achieve more accurate estimates, the current summary provides effect size estimates that should be considered in respect of to individuals exposed adverse thermal conditions.

INTRODUCTION

How the thermal environment affects human response capacity has been the subject of both theoretical speculation and experimental evaluation since before psychology or even physiology became recognized sciences. Implicit evaluations of thermal effects are evident in the site selection of habitation and are influential in the earliest formal military conflicts (Goldman, 2001). More scientifically stringent observations were first generated as a result of practical problems faced by manufacturers whose processes involved the exposure of workers to temperature conditions in which it was uncertain that they could survive. Originally, it was believed that humans could not tolerate levels of thermal stress which exceeded the boiling point of water but an early and adventurous empirical demonstration proved the fallacy of this assumption (Blagden, 1775a, b). The influence of the thermal environment on behavior also played an evident role in patterns of global colonization. For example, Henry Ellis, the then English Governor of Georgia in what was to later become the United States, made a number of adverse comments about the summer heat in that locale which to a degree and for a time served to discourage further European immigration to the area (Ellis, 1758). These and like empirical observations provided the partial foundation of the subsequent theory of environmental determinism. The strict interpretation of this latter theory had it that environmental conditions, among which temperature was conceived as a crucial controlling factor, provided the key to the understanding of essentially all culture and behavior. (e.g., Huntington, 1919; Von Humboldt, 1816).

The impact of this growing knowledge of thermal effects on behavior however was not solely confined to cultural, colonial, or even industrial issues. At the turn of the twentieth century the study of temperature regulation was central to all of physiology, especially following upon Bernard's crucial observation on the "fixity of the interior environment," made in his classic 'Cahier Rouge,' which was primarily concerned with the stability of internal bodily process and especially the stability of temperature itself (and see Hoff, Guillemin, & Guillemin, 1967). This notion of a regulated internal environment was elaborated and articulated most clearly in Cannon's (1932) influential exposition on the concept of

homeostasis. The codification of the notion of controlling feedback systems which underlie bodily system regulation (Hancock, 1980; 1981a) had a fundamental and continuing effect on many fields of study, while exerting its strongest influence on the foundations of physiology, psychology, and the broader areas of cognition and neuroscience.

At no time were the on-going concerns for the practical issue of temperature effects on performance far behind contemporary theoretical developments. For example, in the very earliest years of the twentieth century, the garment industry in Lancashire in England was crucially dependent upon the heat and humidity levels which prevented exposed cotton fibers from drying and breaking during cloth manufacture. How these obligatory heat stress conditions affected workers was a central issue to the British Industrial Fatigue Research Board, (BIFRB; e.g., Vernon & Bedford, 1930), which was itself the precursor to many subsequent world organizations in Public Health and Safety. Practical concerns also drove South African interests in heat and humidity effects since the efficiency of gold mining was crucially dependent upon understanding how workers could acclimatize to and subsequently perform in the exceptionally high heat and humidity levels of the deep mines (cf., Wyndham, 1969). These collective studies showed the advantages of artificial acclimatization procedures on work productivity including beneficial effects on each of physical, psychomotor, and cognitive activities (e.g., Goldman, 2001; Patterson, Taylor, & Amos, 1998; Wyndham, Strydom, Munro, MacPherson, Metz, Schaff, & Schieber, 1964). Nominally, pragmatism also drove thermal research to immoral depths, when investigators under the Nazi regime immersed prisoners in freezing water to determine absolute tolerance times for fatal hypothermia and so putatively aid with the survival of pilot's shot down over the North Sea (see Burton & Edholm, 1955). For a plethora of reasons such experiments have been repudiated as science and discredited as evaluation procedures (and see Hancock, 2003). More recently, human exploration of outer space triggered further efforts to understand how performance capacity varied under thermal extremes, especially in the circumstances experienced during atmospheric re-entry, which it was thought at one time might be too extreme for human astronauts to tolerate (Blockley & Lyman, 1950, 1951).

The experimental approach with the longest history in understanding thermal effects on performance capability is that which treats stress as a property of the environment itself. Derived originally from engineering approaches to the question of material stress exposure, these experiments proceed by varying some characteristic of the thermal environment and measure subsequent effects on reflections of response capacity such as comfort (Fanger, 1967), psycho-motor and cognitive performance (Poulton, 1970; Poulton & Kerslake, 1965), or simple survival (Taylor, 1948). In recent decades, this approach has diminished as risk-averse human participant Review Boards have become ever less willing to permit the exposure of individuals to potentially damaging, and extreme, environmental conditions. Despite this trend, this form of research is still practiced, particularly in special circumstances such as the imperative demands for military test and evaluation (Harris, Hancock, & Harris, 2005; Johnson & Kobrick, 2001). These evolutions in experimental strategies are however, evident in our meta-analytic observations, which show the diminution of these environmental exposures that have been reported over the more recent decades. Contemporary research has most evidently been influenced by Lazarus's influential notion of stress as a transactional process (e.g., Lazarus & Folkman, 1984). Such studies are much more focused upon the process of coping and appraisal in which the stress is an emergent property of the interaction between the individual and the ambient condition. Indeed, recent research has established that perception of heat-induced pain is reduced by relatively high arousal derived from fear induction (Rhudy & Meagher, 2003). In such contexts, however, thermal stress remains an important theoretical and practical concern. There are a number of reasons why thermal stress still occupies this crucial role.

It is a supportable proposition that thermal variation is the modal form of stress faced by all living organisms including human beings. Insights derived from the action of thermal stress therefore serve to inform studies on all forms of general stress effect. This assertion is supported by a number of evident facts. First, temperature is a property of the environment but also a property of the individual thus, unlike most other environmental sources of disturbance such as noise and vibration; there is a direct analog of

the environmental stress already resident in the organism itself. This association between an exogenously measurable characteristic of the environment and an internal representation of the same characteristic allows us to derive clear causal linkages which unfortunately are much more complex in the case of other sources of stress. Second, we know that like many other organisms, humans oscillate in response capacity across time of day (Kleitman, 1939; 1963). This diurnal variation is tied to intrinsic circadian rhythms and the prime physiological indicator of circadian phase is core body temperature (Aschoff, 1984). Thus, as we seek to understand temperature effects, we already have a strong a priori foundation for expecting a direct relationship between body temperature and performance. Third, thermal conditions are stressful to humans in both their excess and their insufficiency. Therefore, unlike comparable environmental sources of disturbance, such as vibration, the low end of the ratio scale of the physical stress is not a comfortable but rather a fatal condition. Thus, extreme cold eventually proves as fatal as extreme heat. This, of course, is due to the thermal dependence of the biochemical platform upon which life itself is erected (cf., Prosser & Nelson, 1981). A final, pragmatic advantage in studying thermal effects lies simply in the number of experimental studies conducted on thermal influences as compared to other sources of environmental stress. As the current meta-analysis shows, there is a substantive existing literature and this also helps establish the veridical pattern of effects more effectively.

As a consequence of these and allied reasons, we examined the effects of thermal stress on performance response capacity. Since our concern is primarily with performance, we have not provide quantitative analysis of the effects of thermal stress on the perception of comfort or physiological or neurophysiological functioning directly, but have focused directly on the influences that such changes have on psychomotor, perceptual, and cognitive capacities. Issues like comfort and the effect of affective reaction on pain perception (e.g., see Rhudy & Meagher, 2003) are, of course, pragmatically important for issues such as indoor occupancy but such effects are addressed elsewhere (see Fanger, 1967; Hancock, 2006). Rather, we report on thermal influences on task performance which contain a preponderance of demands on the information processing capacities of the exposed individual, although some reports do

include at least an element of required physical effort. The purpose of the current meta-analytic review was to understand these performance effects in detail.

ANALYTICAL METHOD

Literature Accumulation

A literature search was conducted using the PsycINFO, MEDLINE, and Dissertation Abstracts International databases, using the following initial key words: 'thermal', 'temperature', 'hot', 'cold', and 'heat'. After a preliminary listing of articles was obtained, references from the obtained journals were examined and article citations were also input into the Science Citation Index. In a concurrent process, subject matter experts were consulted for articles that had not been identified by the formal search procedure (and see Pilcher, Nadler, & Busch, 2002). All articles so referenced were then searched for additional reference information. When these processes no longer yielded new citations we compiled our final listing of articles. This process resulted in identification of 291 articles, reports, dissertations, and theses published between 1925 and 2004. Of these, 49 papers were identified, containing 57 primary studies, which met six selection criteria for inclusion. These specific papers are identified in the present reference list by an asterisk appearing in front of the first author (American Psychological Association, 2001).

Identified Criteria for Study Inclusion

All studies were inspected to ensure that they fulfilled the following six criteria for inclusion in the meta-analysis:

1) Each study reported an empirical examination of thermal stress in which the experimental manipulation involved the explicit application of either a heat or cold stress.

2) A room temperature control group was employed for comparison purposes.

3) Sufficient information was provided regarding environmental temperature conditions in order to be able to generate a Wet Bulb Globe Temperature (WBGT) index value.

4) Temperature exposures had to be whole body air exposures (e.g., not partial body exposures or water immersion, see e.g., Goodman, Hancock, Runnings, & Brown, 1984).

5) Each study had to report on at least one type of performance measure (e.g., accuracy or response time in marksmanship, memory, tracking, detection tasks, etc.). Studies using only comfort votes, physiological, or subjective measures alone were thus excluded from the present analysis.

6) Sufficient information regarding performance measures had to be provided to determine effect size estimates.

It is important to note that rejecting many primary studies in a meta-analysis is a common occurrence and is necessary to ensure meaningful results when combining effect sizes across studies. In addition, and as is often the case (e.g., see Pilcher, Nadler, & Busch, 2002), the modal reason for exclusion of a study was the failure to provide sufficient data for the calculation of such effect sizes.

Identification of the Thermal Conditions In Each Study

Studies included in the meta-analysis had to report the respective aspects of the environmental conditions sufficient to directly represent the environment in the Wet Bulb Globe Temperature (WBGT) index values or to provide enough information to subsequently derive such WBGT values. WBGT was used because it provides a composite measure of physical values of air temperature and humidity and has become the accepted Interantional index of thermal conditions (Parsons, 1993, 1995; Yaglou & Minard, 1957). Studies reporting the environmental temperature in the form of the Effective Temperature (ET) index (Equation 1; Brief & Confer, 1971), Dry Bulb and Relative Humidity (Equation 2), or Dry Bulb alone (computed as a range value using Equation 2 with minimum and maximum Relative Humidity levels) enabled acceptable conversion to WBGT values. In cases where Dry Bulb and Wet Bulb

temperatures alone were provided, these were used to estimate relative humidity using equations provided by the National Weather service (Southern Region Headquarters National Weather Service, 2005).

WBGT = (ET-13.1)/0.823 ---------- (1)

WBGT = 0.567(DB) + 0.393(RH) +3.94 ---------- (2)

The Calculation of Effect Size

Effect sizes used for the current study were the standardized mean differences between the experimental and the control conditions, often referred to as Hedges' g (Hedges & Olkin, 1985, p. 78; see also Hedges, Shymansky, & Woodworth, 1989). Many researchers may be more familiar with Cohen's d (Cohen, 1988), which is conceptually identical to Hedges' g but statistically different (see Hunter & Schmidt, 2004). When means and standard deviations were available the effect size was calculated using equation 3,

$$g = \frac{\left(\overline{X}_E - \overline{X}_C\right)}{s} \quad \text{---------- (3)}$$

Where $\overline{X}_E$ = mean of the experimental condition, $\overline{X}_C$ = mean of the control condition, and s = standard deviation (for the control condition in within-participants designs; pooled standard deviation for between-participants designs). When the means and/or standard deviations were not available, effect sizes were obtained from inferential statistics (e.g., *t*-tests) or sums of squares/mean squares (e.g., from ANOVA tables) using equations found in Lipsey and Wilson (2001) and Hedges, Shymansky, and Woodworth (1989).

In calculating each effect size, the sign was controlled to ensure that a positive score represented improvement in performance in the experimental group relative to the control group, whereas a negative score indicated performance impairment. The mean weighted effect sizes were computed by weighting

each effect size by the reciprocal of its variance, using procedures described in Hedges and Olkin (1985) and with variance formulae in Morris and DeShon (2002). Prior to the weighting procedure, each effect size was adjusted for statistical bias using established procedures (Hedges & Olkin, 1985), and were also adjusted to a common standard deviation (i.e., standard deviations of the experimental and control groups rather than the standard deviation of differences) using procedures described by Morris and DeShon (2002). To correct the g scores for statistical bias, which decreases the accuracy of the estimates particularly in cases where sample sizes are small, an adjustment was preformed. This adjustment (see Hedges & Olkin, 1985, pp. 78-81) provides an unbiased d.

$$d = g * \left(1 - \frac{3}{4 * N - 9}\right) \text{ ---------- (4)}$$

The need for this correction comes from the statistical bias associated with d as an estimate of the population parameter (δ). The expectation for d is given by E(d)=δ/J(N-2), where J(m)= 1 – [3/(4m-1)]. Thus, the expected value of d is not the population parameter δ but is proportional to it (and see Hedges & Olkin, 1985, p.79).

It is important to note that most of the studies used in the current analysis reported multiple g-scores for multiple temperatures, tasks, or durations; for a total of 528 effect sizes (181 for cold and 347 for heat). Many of the studies included multiple effect sizes derived from common participant samples, so these estimates are not independent of one another. Violations of the independence assumption can lead to underestimation of the variance due to sampling error (Cheung & Chang, 2004; Martinussen & Bjornstad, 1999). To avoid such violations we averaged the effect sizes within studies prior to estimating means and variances (see Lipsey & Wilson, 2001). Note, however, that in the moderator analyses this averaging procedure was conducted within each level of the respective moderator variable. For instance, if a particular study contributed an effect size for each of two task categories, these effect sizes were not averaged for that moderator analysis but were included in their respective groups. Since the levels of the

moderator variables were never formally (i.e., statistically) compared, this does not represent a violation of the independence assumption. However, for the global analysis and other moderator analyses not including task, these effect sizes were averaged. Thus, the number of studies within each level of a moderator level is not necessarily the sum of the total number of studies used in the global analysis. Such problems associated with the independence assumption are intrinsic to meta-analyses generally and not simply a characteristic of the current work.

In addition to the weighted mean effect size, two variance estimates were computed: variability due to sampling error (s_e^2) and variability of the effect sizes (s_g^2). These values were used to estimate the variability due to differences in the population effect sizes, δ (s_δ^2). Thus,

$$s_\delta^2 = s_g^2 - s_e^2 \quad \text{---------- (5)}$$

A large s_δ^2 indicates that there is variability among the observed effect sizes that cannot be accounted for by sampling error and that there are likely to be one or more variables which additionally moderate the magnitude of the effect in question (see Hunter & Schmidt, 2004, pg. 288). If all of the variance in the effect sizes were accounted for by sampling error, then $s_\delta^2 = 0$. The 95% confidence intervals reported were computed using the standard deviation corresponding to s_e^2. In some cases the estimate of the variance of the effect sizes can be less than the estimate of sampling error variance. This is conceptually similar to a treatment mean square in an F ratio being less than the mean square error. Indeed, Hedges and Olkin (1985) recommend a statistical test using these variances to test hypotheses regarding any moderator variables. We have not adopted this approach as this test can be biased, particularly when the sample size is small (Hunter & Schmidt, 2004) and our purpose is to transcend the limitations of significance tests by application of meta-analytic techniques. Instead we use the '75% rule' guideline recommended by Hunter and Schmidt (2004). According to this guideline, if 75% of the observed variance (s_g^2) is due to sampling error (s_e^2), it is likely that most of the other 25% of the variance is also artifactual. Cases in which this condition is not met can be interpreted as instances in which the residual

variance (i.e., s^2_δ, the variance not accounted for by sampling error) are likely to be 'real' variance due to differences in conditions across studies. Hunter and Schmidt (2004) note that the 75% rule is as at least as powerful as more formal significance tests for homogeneity, and with small numbers of effect sizes is actually more powerful. Perusal of the tables reported here indicates that, in general, application of the 75% rule results in the conclusion that there is substantial residual variance (i.e., $s^2_e / s^2_g < 0.75$). Thus, in general, the results reported in the present tables indicate that there are likely moderator variables unaccounted for in the present analyses. This underscores the heterogeneity across studies of the effect of thermal stress on human performance.

RESULTS

The 181 effect sizes for cold and 347 effect sizes for heat were obtained from a total of 57 primary studies as specified by an asterisk notation in the present reference list. Several moderator analyses were computed based on the temperature range, the tasks which were employed, the type of dependent measure(s) used, and the duration of the exposure to the thermal stress. These moderator analyses were accomplished separately and, when possible, hierarchically (e.g., by task within each temperature range). The latter strategy is analogous to an analysis of interactions among variables. A limitation of such analyses is that division of effect sizes into separate categories quickly reduces the absolute number of these effect sizes upon which the means and variances can be estimated. Thus, there are instances in which some levels of a moderator variable contain relatively few studies. Such estimates, as we note, should therefore be interpreted with the appropriate caution. However, these cases reveal the paucity of studies examining particular combinations of relevant variables and therefore point toward the need for further experimentation to reliably identify these specific combinatorial effects.

Outlier Analysis

The residual variances observed in these respective analyses tend to be quite large relative to the error variances. One possible reason for this is the presence of outliers or influential data points. Prior to

present overall analysis, one study, that of Beshir, El-Sabagh, and El-Nawawi (1981) was removed because the effect sizes were clear outliers (the effect sizes for their heat stress conditions were g= -12.76 and g=-15.62 for the 26°C and 30°C WBGT conditions, respectively; and see Pilcher et al., 2002). The potential presence of other, more subtly influential data points was tested formally using techniques developed by Huffcutt and Arthur (1995). They introduced a measure conceptually similar to DFFITS (the difference in fit value; the change in the predicted value from the exclusion of a particular case) that accounted for the influence of sample size associated with each effect size. This measure, the sample-adjusted meta-analytic deviancy statistic (SAMD), is therefore an index of the degree to which an individual effect size influences the overall mean effect size. This procedure was employed in the analysis here.

Global Effects of Heat and Cold

As Table 1 shows, the overall effect size of temperature on performance was g=-.34. Consequently, performance under thermal stress proved on average to be approximately one-third of a standard deviation or about 11% worse than performance at a comparative thermoneutral temperature. This unsurprising finding confirms the expectation that thermal stress adversely affects human information-processing and psycho-motor capacities. As the confidence interval of this value excludes zero we are confident as to the reality of this effect. Table 1 also divides the present results into the different respective effect sizes for heat and cold. The effect sizes for each form of thermal stress were comparable, with g= -.29 for heat and g= -.26 for cold. Thus both heat stress and cold stress each exert deleterious influences and individually, neither has a zero level effect within its confidence interval.

Outlier Analysis. Analysis of the entire data set revealed eight outliers which represented influential datapoints among the fifty-seven total studies. These were the studies by Beshir, El-Sabagh, and El-Nawawi, (1981), Lockhart and Kiess (1971), Mayo (1955), Lockhart (1968), van Orden, Benoit, and Osga (1996), Pepler and Warner (1968), Blockley and Lyman (1950), and Peccolo (1962). Exclusion

of these studies from the meta-analysis resulted in a larger overall negative effect (g=-.47, -.56<δ<-.38) with slightly larger error variance (s_e^2 =.10) and smaller residual variance (s_δ^2 =.39). However, the pattern evident in these collective results is comparable whether one includes or excludes these outliers.

<Insert Table 1 here>

Intensity Effects

In a subsequent hierarchical analysis a more detailed examination was conducted which derived a finer discrimination with respect to the range of temperatures. The first division was performed on the heat stress effects and was a straight division of effects above or below the threshold of 85°F ET. This boundary forms the threshold of Lind's (1963) "prescriptive zone" which was integrated into a more comprehensive description of heat effects by Hancock and Vercruyssen (1988) as the threshold of the "zone of thermal tolerance." The value of 85°F ET represents the thermal condition in which the body begins the process of obligatory heat storage. In this circumstance though the individual is dissipating heat at their maximal capacity they shall experience an uncompensable dynamic increase in core body temperature. Consequently, this value has been taken as a crucial thermal threshold (Hancock, 1984, 1986a). As is evident from Table 1, the *average* effect size of these two divisions is very similar (i.e., -.25 vs -.27). However, what is evident is that there is almost threefold greater variability in those studies which make up the group above 85°F ET (i.e., s_g^2 = 0.67) versus those below (i.e., s_g^2 = 0.24). The conclusion to be drawn from this observation is that there are additional influences in the former group. We believe that exposure time plays a very important role here. Given this three-fold difference in effect-size variability we can confirm that 85°F (29.4°C) ET, and its WBGT equivalent (87.4°F; 30.8°C) remains a useful theoretical and practical threshold to parse heat stress effects.

Outlier Analysis. An outlier analysis of this category identified one outlier in the ET>85°F category (i.e., Blockley & Lyman, 1950). For the ET<85°F category four outliers were identified (i.e.,

Hygge & Knez, 2001; Mayo, 1955; Allen & Fischer, 1978; Peccolo, 1962). When these outliers were removed and the meta-analysis recalculated using the remaining studies, a somewhat different pattern of results were observed compared to the original analysis. Higher temperatures (ET>85°F) resulted in larger effects ($\overline{g}$ = -.36, s_e^2 =.09, s_δ^2=.40; 95% CI:-.46<δ<-.25) than when temperatures were below this threshold ($\overline{g}$ =-.17, s_e^2 =.03 s_δ^2=.06; 95% CI: -.27<δ<-.07). The confidence intervals associated with each temperature category in the re-analysis now overlap only slightly. In contrast, with the outliers still included, the effect sizes for the two categories were almost identical (see Table 1). Consequently, conclusions that can be drawn depend upon whether the five studies are included in the moderator analysis. However, one possible reason there were outliers in the ET<85°F might be sample size. Three of the four studies in that category (the exception was the Allen & Fischer, 1978 report) had sample sizes much larger than those in used in the other studies in that category. It has been argued that removing outliers prior to analysis renders the mean effect size more 'representative' of the population of studies. Taken together, the results indicate that Lind's prescriptive zone differentiates the effectiveness of thermal stress in impairing performance, but that other variables likely exist that moderate this division. That is, whether the 85°F ET threshold is a useful differentiation depends on the presence of certain other conditions.

Empirical Division of Temperature Range.

Table 1 further parses the impact of heat. However, this latter differentiation is on the basis of an empirical rather than theoretical division. This empirical grouping gives four naturally distinct ranges starting first below 78.3°F (25.7°C) ET; then from that 78.3°F to 85.0°F (29.4°C) ET; from 85.0°F (29.4°C) ET to 95.4°F (35.2°C) ET; and finally a range of 95.4°F (35.2°C) ET and above (see Table 2 for a listing of these studies). These values were empirically driven by taking the median division of the studies below 85ºF (29.4°C) ET and the median temperature used in studies above 85ºF (29.4°C) ET. These the lowest intensity, the effect size is large, g= -0.97, while that of the most intense temperature

exposures the effect size is moderate, g= -.39. However, as observed in Table 1, the variability of the effect size in the highest heat stress group is now nearly *seventeen times* higher than that of the lowest group. The error variance for the highest heat stress group is comparable to that of two of the other three temperature ranges, including the lowest. Thus, factors other than error variance appear to be driving the differences in variability. With the exception of the lowest temperature range, it is apparent that the effect size variation sequentially increases across the three remaining categories. This gives rise to the proposition that performance is relatively stable over much of the temperature range but exhibits radical variation at the highest extreme. This observation, that performance stability is a crucial indicator of incipient failure is a central characteristic of the extended-U theory of stress and performance capacity (Hancock & Warm, 1989). Table 1 also illustrates similar trends for the influence of cold. Above a value of approximately 52°F (11.1°C), there is a small to moderate increment in performance which in respect of previous observations on transient effects is not surprising (Poulton, 1976). However, below this value there is a large deleterious effect. Thus, as with heat we have a clear stress intensity effect which, while not unexpected, is important to confirm here.

<Insert Table 2 here>

Outlier Analysis: Cold. At temperatures below 52°F, two outliers were identified (Lockhart & Kiess, 1971; van Orden, Benoit, & Osga, 1996). Removal of these outliers resulted in a substantial attenuation of the mean effect size ($\overline{g}$ = -.56, 95% CI: -.79<δ<-.33) and the residual variance (s_δ^2=1.00), although the error variance was similar (s_e^2=.14). At temperatures above 52°F three outliers were identified (Lockhart, 1968; Pepler & Warner, 1968; Sharma & Panwar, 1987). Inspection of the data indicated that the effect sizes for these studies were much larger than for the others in the same category. Removal of these studies and re-analysis yielded a much smaller effect size ($\overline{g}$ = -.10, 95% CI: -.29<δ<.10) and residual (s_δ^2=.08), with similar level of error variance (s_e^2=.06). Indeed, applying Hunter and Schmidt's 75% rule, we conclude that the six remaining studies are homogenous. However,

the confidence interval associated with this mean does include zero. Thus, the analysis without the influential data points indicates a null effect, while inclusion of these studies indicates a small to medium *positive* effect on performance. This is not due to changes in statistical power; because the error variances in the two analyses are similar (slightly smaller in the analysis with the outliers removed). Therefore, six of the nine studies in this category indicate a reliable null effect of moderate cold on performance.

Task Type

In Table 1 we have also parsed the temperature effects into three sub-divisions, based upon the differentiation of information-processing stages (cf., Lachman, Lachman, & Butterfield, 1979; Wickens & Hollands, 2000; and see also Pilcher et al., 2002). This tripartite differentiation splits effects into perceptual, cognitive, and psychomotor response capacities which is important since change in response capacity under thermal stress is not equivalent in each case (see Grether, 1973; Hancock, 1982; Ramsey, 1995; Ramsey & Morrissey, 1978). These do not represent a specific model of human information processing, but rather are general task performance categories for the appropriate differentiation of varying task demand. There are detrimental effects for thermal stress on all three components with the highest impact being on perception, the next highest being on psychomotor response, while the smallest effect was observed for the cognitive task category. The variability of the effect sizes is again greater in the higher impact categories of perception and psychomotor response.

Performance Measures

In analyzing performance, a primary division of the dependent variable is often split between the speed of a response and the accuracy of that response. In many circumstances speed is traded for accuracy and vice versa (Fitts, 1954). Although the number of effect sizes for accuracy was almost double that for response time, there was sufficient evidence to provide a stable representation of each. As is evident in Table 1, thermal stress generates substantial performance degradation on both response speed and

response accuracy, with the larger effect being on speed of response. This outcome argues against a speed-accuracy trade-off being responsible for results previously observed which, if present, represent a strategic change rather than an absolute reduction in response capacity. The variability between the effect sizes which compose these separate influences on response time and response accuracy is evidently high, although the error variances are comparable. With a three-fold difference it is clear that response-time effect-sizes are composed of individual effects of much greater variation across studies, indicating the likely influence of moderator variables on response time that are not evident for accuracy.

Duration of Exposure

One crucial factor which permeates all stress effects is exposure time. To evaluate its effects we subdivided studies according to duration of exposure and these results are shown at the bottom of Table 1. It is evident here that duration does exert a systematic effect and if we consider the first three categories up to three hours in duration, we can see that there is a sequential increase in effect size with increasing exposure time. Performance is affected adversely to an ever greater degree as time-on-task proceeds. While this effect is relatively stable, the effect size variability that accompanies these temporal effects also increases. This pattern is consistent with the expectation of a time/intensity effect in research on stress in general and the degree of variation also being a mark of progressive instability (see Hancock & Warm, 1989). However, this pattern is broken in the final category of exposure beyond three hours. As indicated in Table 1, durations above three hours prove essentially as benign as exposures below one hour and the error variances are similar. This outcome is however, an artifact. As is evident from perusal of the original studies, only the lower temperature ranges (i.e., less intense exposure) are tolerable for this extended period of time. How these duration effects are modulated by acclimatization is surely an important issue. However, as indicated in Table 3, so few studies specifically included an acclimatization regimen that there is insufficient present information to perform a meta-analysis on such effects, although this in itself indicates an important direction for future research.

<Insert Table 3 here>

In the present findings, there is an overriding interaction between exposure time and tolerable exposure temperature. This is not immediately evident since effect sizes are initially the equivalent of main effects. The time by intensity outcome is fundamentally an interaction and this issue is re-examined when we later consider multi-factor effect-sizes below. Results from the present meta-analysis serve to confirm duration of exposure effects and this quantitative summation is consistent with more qualitative, previous observations on heat stress (Hancock, 1981b; 1982; 1986b; Hancock & Vasmatzidis, 1998; Ramsey, 1995) and some similar observations on stress in general (e.g., Hancock & Desmond, 2001; Hockey, Gaillard, & Coles, 1986).

Outlier Analysis. The outlier analyses for the duration categories revealed influential data points in each category. Two outliers were observed in the 'less than one hour' category (Lockhart, 1968; van Orden, Benoit, & Osga, 1996), and in the 1-2 hour category (Lockhart & Kiess, 1971; Blockley & Lyman, 1950). Five outliers were observed in the 2-3 hour category (Beshir, El-Sabagh, & El-Nawawi, 1981; Lockhart & Kiess, 1971; Wyon, 1969; Bell & Provins, 1963; Griffiths & Boyce, 1971), and three outliers were observed in the longest duration category (Lockhart & Kiess, 1971; Pepler & Warner, 1968; and Mayo, 1955). Re-analysis after removal of these influential data points resulted in much smaller s_g^2 and a different pattern of effect sizes across duration categories. These are shown in Table 4. The mean effect sizes in the two lower duration categories and in the highest duration category increased, while the effect in the 2-3 hour category decreased. Examination of the relevant confidence intervals suggests that there is a substantial increase in effect from less than one hour to 1-2 hours and that beyond this the effect of duration varies as a function of other moderating variables and sampling error.

<Insert Table 4 here>

Hierarchical Analysis: Task Category and Performance Measure

The findings in regard to task category and performance measures were analyzed hierarchically by heat and cold effects individually and these results are given in Table 5. As is clear, heat stress detrimentally influences all of the three performance categories. Most affected by heat stress are perceptual capabilities followed by motor responses with cognitive attributes least affected but still subject to decrement. In this, the heat stress findings replicate the overall thermal effects. For cold, this pattern is again confirmed with the motor and perceptual capacities being affected to the greatest degree. However, the pattern for cold is not exactly equivalent to that for heat. Perceptual abilities are most affected but the level of that effect is greater (i.e., -1.13 vs. -0.78). Motor response represents the intermediate category, with similar levels of decrement in heat and cold (i.e., -0.42 vs. -0.31). However, a clear difference is evident in the area of cognition (see Table 6). The effect of cold on cognition is an *increment* in performance capacity compared to the decrement for heat (i.e., +0.41 vs. -0.23). This may be due to the transient warming effects of cold on core body temperature as we discuss later.

<Insert Tables 5 and 6 here>

As performance is differentiated by task, so the main results can be distinguished by their effects on response time and accuracy. In Table 4, we have subdivided the heat and cold effects into the two performance elements. Heat is deleterious to both speed and accuracy of performance. Although the variation in effect size distribution is greater for response time, the error variances are comparable and thus the outcome is sufficiently stable to confirm these respective influences. The cold influences are much less deleterious in general. As again is clear, cold stress reduces response time capability but only by a relatively small degree. The most interesting finding is that cold exerts almost no effect on accuracy. Here, the zero (no effect) is contained within the 95% confidence intervals for both response time and accuracy in the cold condition. However, the large variance in g for the cold conditions indicates that the small effect sizes may be due to the influence of other moderating variables. The general conclusion, that

heat is somewhat more damaging then cold is in line with physiological accounts of thermal stress effects in general (Hancock, 1986b).

Joint Effects of Intensity and Task Type

Subdivision of the respective stress effects has two conflicting tendencies. First, sub-division allows us to look much more informatively at the influences of multiple factors. As one cannot truly understand performance without both speed and accuracy measures, so one cannot understand many stress effects without grasping the combined impact of the crucial factors of time and intensity. Unfortunately, each subdivision of the effect matrix reduces the number of effects involved and thus the reliability of the observed outcome. Soon, one is down to a single effect from a single study or even worse, no effect reported in a particular category at all. This tendency is illustrated in Table 7 in which we have parsed the information by general task-demand category and heat stress level above and below the 85°F (29.4°C), E.T. (see Hancock & Vercruyssen, 1988; Lind, 1963). This results in a very different pattern for heat effects contingent upon the type of task being performed. Unfortunately perceptual capacity could not be examined due to the problem of the restricted number of effects as mentioned earlier. Fortunately, this problem does not hold for the other task-demand components. Motor responses to heat prove very systematic. Above 85ºF (29.4°C), E.T., there is a moderately strong decrement and the error variance confidence interval excludes a zero effect. Conversely, motor response seems to be unaffected below the 85ºF, E.T. threshold, as shown by the confidence interval, in which the range of g scores brackets the zero effect. The most interesting pattern however, is evident for cognitive activity. While there is essentially no effect upon cognitive tasks over the 85ºF (29.4°C) ET threshold, there is a reasonably strong negative impact on cognitive performance below the 85ºF (29.4°C) ET threshold. It would appear that studies employing less intense heat stress actually lead to a greater cognitive task decrement. This bifurcation of effects across the tasks confirms the earlier observations of 85°F (29.4°C) E.T. as a crucial heat stress threshold.

<Insert Table 7 here>

Outlier Analysis.

Perceptual Tasks. Analysis of effect size differences between different divisions of heat within each task category indicated that for motor tasks no outliers/influential data points were observed. Among studies using perceptual tasks at temperatures above 85°F ET, two outliers were identified (i.e., Hocking, Silberstein, Lau, Stough, & Roberts, 2001; Courtright, 1976). Removal of these resulted in a substantially smaller mean effect size ($\overline{g}$ = -.29, 95% CI: -.43<δ<-.15), as well as smaller sampling error variance (s_e^2 =.08) and residual variance (s_δ^2=.16).

Cognitive Tasks. For studies with cognitive tasks and exposures below 85°F ET, three outliers were identified (Peccolo, 1962; Mayo, 1955; Hygge & Knez, 2001). After removal of these studies, subsequent re-analysis revealed a much larger mean effect size ($\overline{g}$ = -.60, 95% CI: -.79<δ<-.42) the residual variance was smaller (s_δ^2=.32). The error variance was comparable to that associated with the analysis with the original data points included (s_e^2 =.08). For the group of studies in which cognitive tasks were used at exposures greater than 85°F ET, two outliers were identified (Bell, 1978; Blockley & Lyman, 1950). However, removal of these did not substantively change the mean or the conclusion of null effects, although the residual variance is smaller with these studies removed.

Psychomotor Tasks. Evaluation of motor tasks at temperatures above 85°F ET revealed two outliers (i.e., Teichner & Wehrkamp, 1954; Epstein, Keren, Moisseiev, Gasko, & Yachin, 1980). However removal of these from analysis did not substantively change the mean effect size or the error variance, although the residual variance was cut in half. No outliers were observed at temperatures below 85°F ET, possibly due to the limited number of studies and the absence of any residual variance (see Table 4).

Performance Measures Within Each Heat Stress Category

Since it is evident that response time and accuracy are differentially affected by heat stress, we have again sub-divided these influences through the parsing of the data by the heat threshold we have previously identified above. These data are presented in Table 6, and show that accuracy is ubiquitously affected by heat stress, regardless of whether the heat is above or below the 85°F (29.4°C), E.T. threshold. The outcome for response time is rather different. Below the threshold value there is essentially no effect for heat on response time and by implication heat stress degradation in this region is due fundamentally to a change in response accuracy. However, as one enters the range of heat stress that cannot be compensated for by normal physiological response (Hancock, 1982; Lind, 1963; Houghton & Yagloglou, 1923), the response time in relation to imposed task demands increases and thus performance becomes significantly worse as both aspects of response capacity are affected.

Outlier Analysis. Outlier analyses of accuracy and speed effects within each temperature category revealed conclusions substantively different when these influential data points were removed from the analysis. Thus, at temperatures above 85°F ET, two outliers were observed for the accuracy analysis (Blockley & Lyman, 1959; Epstein, Keren, Moisseiev, Gasko, & Yachin, 1980), and removal of these yielded similar values of mean effect size and error variance for the ET>85°F category ($\overline{g}$ = -.38, s_e^2 =.09, 95% CI: -.50<δ<-.26), and less residual variance (s_δ^2=.34). For the speed analysis four outliers were identified (Hocking, Silberstein, Lau, Stough, & Roberts, 2001; Courtright, 1976; Epstein, Keren, Moisseiev, Gasko, & Yachin, 1980; Pepler & Warner, 1968). However, the analysis of speed in the same heat category indicated that removal of the outliers resulted in a negligible effect of $\overline{g}$ =-.08. Even with an associated smaller sampling error variance (s_e^2 =.06), the confidence interval still included zero (95% CI: -.21<δ<.06). However, the residual variance was over four times greater than the error variance (s_δ^2=.27).

For studies utilizing exposures below 85°F ET three outliers were identified for accuracy (Peccolo, 1962; Hygge & Knez, 2001; Pepler & Warner, 1968) and one for speed (Holmberg & Wyon, 1969). Removal of these from the analysis resulted in a smaller effect for accuracy ($\overline{g}$ = -.18, 95% CI: -.29<δ<-.06) with similar error variance (s_e^2 =.03) but smaller residual variance (s_δ^2=.08). By contrast, removal of one outlier for the speed analysis changed the outcome of the analysis from a null effect observed when the Holmberg and Wyon (1969) study was included to a small negative effect ($\overline{g}$ = -.19, s_e^2 =.02, 95% CI: -.31<δ<-.06) with less residual variance (s_δ^2=.03) when that study was excluded. Thus, accuracy is impaired more at temperatures above 85°F ET, while impairment of speed occurs more at temperatures below 85°F ET. While this is not a true speed-accuracy tradeoff, it does reflect differences in the effect of thermal stress dependent upon the order of performance examined.

Joint Effects of Intensity and Duration

In Table 8, we have parsed the data, this time according to combination of thermal level and exposure duration. While this division promises to be most informative, it does reduce the number of effect sizes to a problematic level in some of the resultant cells. In general, we expected to see a time by intensity interaction such that the longer exposures at the higher thermal stress levels prove to be the most deleterious. For heat stress we do *not* see this pattern. Rather, it is the shorter exposures which result in greater degradation. This is surprising but may reflect Poulton's (1976) invocation of an acclimation factor. Even the variability of effect size factor is not an explanation for this effect. This observation has important implications for length of duty in heat stress as we will discuss later. This temporal factor, also plays a role in activities such as mission planning and shift work considerations. Cold provides a somewhat clearer picture. Above approximately 52°F (11.1°C), E.T. cold acts to improve performance but as exposure time increases this improvement is attenuated. Below 52ºF (11.1°C) ET, durations of less than two hours are more deleterious than anything over 52°F (11.1°C), E.T., see Table 8. Problematically, estimating the effects of exposures over two hours at lower temperatures was limited by the very small

number of effect sizes. As can be seen, this latter estimate depends on only two effect sizes which themselves each provide very different outcomes. At present, it is probably best to treat this result as unreliable and adopt a conservative approach by considering this condition at least equally as damaging as less than two hour duration and an area which should be one priority for future research.

<Insert Table 8 here>

Outlier Analyses: Heat. Examination of the duration by intensity interaction for heat indicated no outliers among effects greater than 2 hours, although this might be due to the small number of studies in these categories. Outliers were however observed among studies with less than 2 hour durations. Thus, for the ET>85°F category four outliers were identified (Blockley & Lyman, 1950; Razmjou, 1992; Courtright, 1976; Pepler, 1960). Removal of these studies and re-analysis resulted in a smaller mean effect size ($\overline{g}$ = -.20, 95% CI: -.36<δ<-.04) but a comparable sampling error variance (s_e^2 =.10). However, the total variance was much lower (s_g^2=.09), and the residual variance was reduced to zero upon removal of these latter influential data points. This indicates that the apparent heterogeneity in this category is due to four of the 19 total studies of which it is composed.

For the group of studies in which ET<85°F and duration of exposure was less than two hours, one outlier was identified (Hygge & Knez, 2001). Removal of this study reduced the mean effect size ($\overline{g}$ =-.33, 95% CI: -.47<δ<-.19), and resulted in a lower residual variance (s_δ^2=.08) and slightly lower error variance (s_e^2 =.03). Consideration of these mean differences and the high degree of overlap between confidence intervals, the apparently anomalous effect in the original analysis, that temperatures *greater* than 85°F ET induced a *smaller* effect, is likely due to the influence of the outlier studies identified. This is further evidence that the utility of Lind's prescriptive zone in understanding performance effects is dependent on other moderating conditions in addition to duration and temperature of exposure.

Outlier Analyses: Cold. Analysis of the duration and cold temperature category interaction also result in different interpretation of the data. Only two studies employed the combination of exposure temperatures less than 52°F and durations greater than 2 hours. For durations less than 2 hours two outliers were identified (Lockhart & Kiess, 1971; van Orden, Benoit, & Osga, 1996). The samples sizes for these latter studies were both n=20, while the other ten studies in that category all had n<=12. Removal of these studies and re-analysis yielded a smaller mean effect size ($\overline{g}$ = -.54, 95% CI: -.77<δ<-.31) associated with substantially smaller residual variance (s_δ^2=1.00). Sampling error variance was similar, however (s_e^2 =.14). The interpretation therefore did not change with removal of the outliers, except that the new magnitude is in the range of medium rather than large effect.

A different story emerges at temperatures above 52°F, however. At durations greater than two hours one outlier was identified (Pepler & Warner, 1968). Reanalysis following removal of this study resulted in a mean effect of similar magnitude but in the opposite direction ($\overline{g}$ = -.67, 95% CI: -1.06<δ<-.29). The effect size for this study was g=1.33, while those of the other three were g=-1.03, -.75, and .21 respectively. Further, the small number of studies removal of Pepler and Warner (1968) resulted in an increase in error variance from s_e^2 =.02 to s_e^2 =.11. However, there was also a substantial decrease in residual variance from s_δ^2=1.44 to s_δ^2=.11. Thus, an interpretation of heterogeneity of variance with all four studies changes to one of homogeneity with the study by Pepler and Warner (1968) removed.

At durations less than 2 hours the removal of the one outlier, Lockhart (1968), resulted in a much smaller and negative mean effect size, although the confidence interval contained zero ($\overline{g}$ = -.13, 95% CI: --.44<δ<.18). This is unlikely to be due to a change in power since omission of Lockhart (1968) reduced sampling error variance from s_e^2 =.11 to s_e^2 =.08. Residual variance was reduced to zero, indicating that the heterogeneity larger effect size observed in the original analysis is due almost entirely to the Lockhart (1968) study. Inspection of the individual effect sizes reveals the reason for this pattern. For the

Lockhart (1968) study the value of g=2.60, while for the other three studies the effect sizes were g=-.10, -.36, and .03, respectively.

Hierarchical Analysis: Joint Effects of Task and Performance Measure

In Table 9, we have extended the process of decomposition of effects by factors, to its greatest practical degree. The division is by thermal stress into its heat and cold components and within this division have looked at each task category and dependent variable. It is perhaps easiest to deal with these categories sequentially and the first observation is that heat stress deleteriously influences perception through a reduction in response accuracy and an increase in response time. For the cognitive task category, heat reduces response accuracy but has a minimal impact on response time.

<Insert Table 9 here>

In the psychomotor task category, it is clear that heat stress exerts strategic effects. The meta-analysis shows that heat reduces psychomotor accuracy; however, the evidence also shows *facilitation* in response time. The actual effect sizes indicate that the facilitation in speed actually outweighs the decrement in accuracy, however, whether the absolute trade-off of these two causes an overall deficiency or increment is still uncertain, since such trade-offs are contingent upon the character of the specific task at hand. The results for cold stress are also systematic but again sparse in places. The first and obvious thing is that, like heat, cold has a significant decrement on both the speed and accuracy of perception. For cognitive tasks, however, the effect size associated with accuracy was small, with the confidence interval including zero. In contrast, cold temperatures produced evidence of an *increment* in response time for cognitive tasks (i.e., better or faster response time). This is also evidence of global speed-accuracy trade-off and again the absolute effect of this is difficult to determine, being task-specific.

In respect of psychomotor activity, we have a further anomalous finding. The average effect size for accuracy shows an evident *increment* in performance (and see Poulton, 1976). We take this to be a

real effect but one that would be very much interrupted with the onset of shivering. Response time shows a decrement indicating a different form of speed-accuracy tradeoff than observed in heat stress. Thus, under heat stress individuals engaged in psychomotor tasks are faster but less accurate, while the reverse appears to be true in cold environments.

Outlier Analyses: Heat Stress

Perceptual Tasks. Three outliers were observed for the accuracy analysis (Bursill, 1958; Razmjou 1992; Chiles, 1958) and two for the speed analysis (Hocking, Silberstein, Lau, Stough, & Roberts, 2001; Courtright, 1976). With the outliers included speed was negatively impacted more then accuracy but the effects were both negative and there was evidence of heterogeneity in the variability across effect sizes (see Table 6). When these outliers are removed, the mean effect sizes are of similar magnitude (for accuracy: $\overline{g}$ =-.30, 95% CI: -.50<δ<-.09; for speed: $\overline{g}$ =-.20, 95% CI: -.36<δ<-.05) and error variance that was similar for accuracy (s_e^2 =.10) and smaller for speed (s_e^2 =.06). Although the residual variance was substantially smaller in both cases, there was still evidence of heterogeneity in effects sizes for speed (s_δ^2=.26), when applying the 75% rule. In contrast, the residual variance for accuracy (s_δ^2=.01) was only 9% of the total variance, indicating a high degree of consistency in the effect of heat on perceptual performance accuracy (i.e., the effect sizes were homogeneous).

Cognitive Tasks. Five outliers were observed for the accuracy category (Peccolo, 1962; Bell, 1978; Hygge & Knez, 2001; Blockley & Lyman, 1950; and Allen & Fischer, 1978). Four of these ranged in effect from g=-1.16 to -1.71. The exception was Blockley and Lyman (1950), who obtained an effect size of g=2.63. Removal of these outliers and reanalysis yielded a smaller effect ($\overline{g}$ = -.15) with a similar error variance (s_e^2 =.03). The residual variance was cut in half (s_δ^2=.24), but the pattern still indicates heterogeneity in the remaining effect. For speed, no outliers were observed. Thus, the outlier analysis for

studies using cognitive tasks reveals no substantively different conclusions with or without the outliers included with the one exception the analysis for the magnitude of the mean effect size for accuracy.

Psychomotor Tasks. Two outliers were observed in accuracy condition (Beshir, El-Sabagh, & El-Nawawi, 1981; Bell, 1978) and one outlier in the speed condition (Razmjou, 1996). The data given in Table 6 represent analyses in which the Beshir, El-Sabagh, and El-Nawawi (1981) study had already been removed for reasons previously discussed. Removal of the effect size obtained from Bell (1978) did not substantively change the outcome of the analysis ($\overline{g}$=-.66, s_e^2=.13, s_δ^2=.63, 95% CI: -.87<δ<-.45). In regard to speed, removal of the outlier reduced the magnitude of the mean effect size ($\overline{g}$=.25, 95% CI: .04<δ<.45) and it also substantially reduced both the error variance (s_e^2=.04) and the residual variance (s_δ^2=.02). Thus, the remaining four studies are homogeneous with respect to the effect of heat on psychomotor speed.

Cold

Cognitive Tasks. There was one outlier/influential data point with respect to cognitive task accuracy (Pepler & Warner, 1968) and one for speed (van Orden, Benoit, & Osga, 1996). Removal of each outlier resulted in a smaller effect for accuracy ($\overline{g}$ = -.15, 95% CI: -.34<δ<.04) with larger error variance (s_e^2=.09) and slightly smaller residual variance (s_δ^2=.47). Thus, a more precise estimate is obtained by inclusion of the outlier. For speed, the mean effect size ($\overline{g}$ = .55, 95% CI: .21<δ<.89) and error variance (s_e^2=.15) were slightly smaller without the outlier, and the residual variance was reduced by more than half (s_δ^2=.21). However, there was still evidence of heterogeneity, so the substantive interpretation does not change when this outlier is removed.

Psychomotor and Perceptual Tasks. One influential data point was identified for accuracy (Lockhart, 1968). Removal of this study resulted in a smaller effect in the opposite direction, although the 95% CI

included zero, rendering this effect somewhat unreliable ($\bar{g}$ =-.24, 95% CI: -.54<δ<.06). This change is not due to less statistical power without that study, since the error variances of the two analyses were the same (s_e^2 =.09). Further, the observed variance was substantially smaller (s_g^2 =.06), indicating that s_δ^2=0. Thus, the variability among the four remaining studies can be attributed entirely to sampling error (i.e., the effect sizes are homogenous). No outliers were observed for the studies using perceptual tasks.

One outlier was identified for speed (Lockhart & Kiess, 1971), and removal of this study rendered a change of mean effect from $\bar{g}$ =-1.10 to $\bar{g}$ =.06, with zero included in the 95% (-.17<δ<.28). However, the error variance (s_e^2 =.04) and the residual variance (s_δ^2 =.20) were also substantially reduced. Given the small number of studies, regardless of whether the outlier is included or excluded, such results should be interpreted with caution. However, the effect size associated with the Lockhart and Kiess (1971) study (g= -4.35) was substantially larger than that associated with the other three (g=.25, .37, and -.87). Thus, this study obtained results quite different from those of the other studies in the same category.

DISCUSSION

With any meta-analysis, what can be derived from the procedure is a quantitative description of the state of present understanding to the degree that it can be known. This description is an empirical map that can address competing theoretical accounts, at least to the degree that any quantitative data can address a theory for which it was not initially derived to evaluate. Although we are very interested in the various theories that have been developed to account for stress effects in general (e.g., Lazarus & Folkman, 1984; Selye, 1976; Ursin & Eriksen, 2004) and thermal stress in particular (Boregowda, 1999; Enander, 1989; Enander & Hygge, 1990), we wish to focus first upon the different descriptive functions which have sought to link stress to performance efficiency.

There are a number of proposed relationships which described the function which relates the level of stress to performance capacity. The most influential ubiquitous of these is the inverted-U which

dominates the landscape, especially in undergraduate and beginning psychology texts. First derived from experiments on discrimination learning in mice (Yerkes & Dodson, 1908), the general function was given particular impetus by Hebb's (1955) influential affirmation. Despite an on-going litany of criticisms, the inverted-U still inhabits this dominant position and despite its obvious flaws, it continues to be taught as one of the fundamental 'laws' of psychology (and see Hancock & Ganey, 2003). We do not rehearse these various objections here. Instead we explore what the present meta-analytic findings mean for this and the other descriptive relationships that have been proposed.

Although the inverted-U is the dominant descriptive relationship (which we have illustrated only in its general form; see Figure 1a), it is by no means the only one. For example, Näätänen (1973) argued that if the individual could keep their attention focused directly on task-relevant cues and not become distracted or overwhelmed by irrelevant stimuli then there should be no descending arm to the inverted-U. In such circumstances performance should continue to increase as stress grows, but with the rate of gain in performance efficiency decreasing progressively with each sequential increase in the level of imposed stress. This results in the function illustrated in Figure 1(b), as compared to the classic inverted-U. This proposal has, most recently, been elaborated and developed by Gaillard (2005) who argues for the utility of the construct of 'concentration' on distinguishing changes associated with change in stress level. It is also consistent with the findings that emotional states associated with high arousal (particularly fear) reduce the subjective perception of pain, thereby mitigating the effects of such pain-inducing environmental stimuli.

<Insert Figure 1 here>

As an elaboration of these earlier postulations, Hancock and Warm (1989) developed the 'extended-U' description of stress and performance change, which is illustrated graphically in Figure 1c (and see Hancock & Warm, 1989, Figure 2). In this conception, stress in the real-world is conceived as being a tolerable form of interference which resulted in only minor levels of change up to specific

threshold limits. At these thresholds (which vary as a function of the type of behavior measured also changes), the various compensatory mechanisms that act to mitigate stress effects each begin to fail and their respective capacities are rapidly overwhelmed as adaptive response now proves insufficient for full compensation (see Figure 2).

<Insert Figure 2 here>

Thus Hancock and Warm's description shows a central, plateau region where performance is *relatively* stable, bound by regions of evident failure characterized by sudden and radical failure and hence the reference to this description as the 'extended-U.' As is evident in Figure 2, the threshold at which comfort fails is much lower in terms of stress level compared to psychological functioning which itself is affected before the onset of physiological failure. This, more general description, was postulated as to apply to physiological as well as performance degradation, in which there is a formalized relation between the failure etiology of physiological as psychological functions.

Given these disparate descriptions, the general decrement in performance with exposure to thermal stress (i.e., either heat or cold stress), can be interpreted as support for either the inverted-U or extended-U hypotheses. The outcome, however, can be seen as somewhat contradict of Näätänen's assertion, but the post-hoc explanation for such a discrepancy would be that participants failed to focus on task-relevant cues, a proposition that is not immediately testable from any of the studies examined. This situation is always of concern when postulating an unobserved, and sometimes unobservable variable, as the crucial causal factor. However, performance did prove to be relatively stable over much of the tolerable temperature range and then showed radical decrement at the highest extremes. This pattern is particularly consistent with the propositions intrinsic to the extended-U theory of stress and performance and directly contradicts the inverted-U description which explicitly defines a single, optimal point of response efficiency. As we have examined explicitly different orders of performance, the usual post hoc adjustment contingent upon an unidentified level of so-called task "complexity" cannot now save the

inverted-U description as it has done in many previous commentaries (but see Hancock & Ganey, 2003). We also observed that there is, in general, an increased negative impact on performance with increased duration of exposure to thermal stress, although this generalization must be tempered with respect to the specific observations we have made on the longest time intervals evaluated.

For our generalized findings concerning thermal stress the Hancock and Warm model fits with the data very well. However, as analyses get more fine grained it becomes necessary to make more specific inferences in regards to task, dependent measure, and other factors. For instance, we found a great deal of heterogeneity in our lowest heat temperature range with two studies with large effects (-1.71 [Allen & Fischer, 1978] and -1.45 [Peccolo, 1962], respectively) driving the overall negative effect in this category while all the other studies demonstrated small effects around the zero level (0.06 [Wyon, 1969], 0.16 [Pepler & Warner, 1968], & -0.16 [Holmberg & Wyon, 1969]). This indicates a greater need for specificity in the division of studies at this level, to discover the driving moderating forces of these disparate results in what is expected to be relatively homogenous results. Another example in which there was a great deal of disparity was in regards to task effects. Perceptual and psychomotor effects frequently proved much larger than cognitive effects. This also supports the postulation in the Hancock and Warm model of dynamic maximal adaptability in regards to its diagnostic ability. Where Hancock and Warm regarded the immediate task as the proximal source of threat and disturbance, not all tasks are equally stressful. Some tasks, such as perceptual and motor demands and their combination in dual-task situations, tend to be more attention-demanding and therefore fail earlier and to a greater degree than do more cognitive tasks (cf., Vasmatzidis, Schlegel, & Hancock, 2002). While most task types were impacted by heat and cold in a relatively similar manner, cognition actually benefited from cold while being negatively affected by heat. The positive effect of cold on cognition compared to the degradation by heat (i.e., +0.41 vs. -0.23), might perhaps be due to the transient warming effects of cold on core body temperature through the thoracic pooling of blood (and see Enander, 1984, 1986; Poulton, 1976). This is supported by the analysis of a median split in cold temperature ranges that demonstrated a beneficial

effect for slight cold while indicating a large decrement in extreme cold in which such transient warming effects would be eventually offset by long-term cooling. Therefore, it was the nature of the experiments with more studies implementing milder cold stress, which lead to the recorded beneficial effect. This serves to remind us that stress is not always a bad thing. In regards to duration, heat seemed to demonstrate some degree of benefit for increased duration, while cold appeared to not demonstrate a benefit with increased duration. However, this difference must be tempered by knowledge of the actual conditions and recognition that one can acclimate to heat but acclimation to cold is presently thought to be untenable. Demonstrating the influence of other such moderators not examined here, especially adaptation to the stress and to the task (Hancock, 1986b) and how these may play differing roles is a goal for future research. Our final observation is that heat generated a common decrement in both speed of response and in response time, but cold demonstrated only a decrement to response time. However, when heat was divided along Lind's threshold, the lower temperature range showed no decrement for response time.

CONCLUSIONS

Hancock and Warm (1989) argued that the nature of the task itself is a crucial influence on the individual's response to stress. The current analysis supports this contention by showing that the effect of thermal environments on human performance vary as a function of task type (and see Grether, 1973). It is likely, however, that task type interacts with other moderators, particularly exposure duration and intensity, in influencing both the magnitude and direction of the effect of thermal stress (Hancock & Vasmatzidis, 2003; Pilcher, Nadler, & Busch, 2002). Further, even in the hierarchical analyses, the variances within most categories indicated the likely influence of other moderators. In such cases, however, there were insufficient studies to explicate these more complex interactions. Thus, future work should examine how these variables jointly influence performance under thermal stress. If such interactions could be explored and quantified (e.g., Loeb & Jeantheau, 1958), it would permit the

construction of vectors representing the collective impact of the variables on adaptation under stress (see Hancock & Warm, 1989).

PRACTICAL APPLICATIONS

One of the issues involved in using stress research for design is the difficulty in quantifying precisely how a particular stressor influences human performance (Poulton, 1965). The current meta-analytic review represents one step toward a solution to this problem by providing quantitative estimates of the effect of a particular stressor on performance on a variety of task types. It is only a partial solution as it does not address how multiple stressors (e.g., temperature, noise, time pressure, task complexity, etc.) combine to influence response outcome (cf., Hancock & Pierce, 1985). Nor, at the present, does it provide completely reliable estimates of specific combinations of task and environmental characteristics even within one stressor category (e.g., the joint effect of task, temperature range, and dependent measure). Because of still evident limitations on the present empirical database, a further limitation of all meta-analyses is that one can only test for combinations of moderator variables that have been investigated to a sufficient degree. The current meta-analytic review underscores the necessity for further empirical research which is purposefully targeted at elucidating these complex interactions. The mean effect sizes reported here represent the current 'best estimates' of thermal effects on performance and can therefore be used to estimate the degree to which particular task-environment combinations will result in performance decrement or performance increment (e.g., Poulton & Kerslake, 1965). For instance, these results indicate that certain combinations of task and environmental characteristics are more likely to induce performance decrement (e.g., performance on perceptual tasks under cold conditions) than other combinations (e.g., response time in cognitive tasks under heat stress). We can therefore recommend their use as guidelines for practitioners who wish to incorporate the effect of thermal stress on operators into their systems design.

It is not only military personnel who have to work under highly adverse thermal conditions. From the glass blowers of Venice to the crab fishermen of Alaska, many occupations require that workers face a significant thermal hazard in their everyday occupation. Our traditional ergonomic strategies of isolation, augmenting support, or restricting exposure times continue to be applicable. However, as we seek to defend people against the deleterious effects of thermal stress it is crucial to capture a quantitative assessment of just what such threats are composed of. The present meta-analysis provides this knowledge to those who design for, supervise, or operate alongside personal in adverse thermal environments.

REFERENCES

*Allen, M.A., & Fischer, G.J. (1978). Ambient temperature effects on paired associate learning. *Ergonomics, 21*(2), 95-101.

American Psychological Association. (2001). *Publication Manual of the American Psychological Association*. Washington, D.C.: American Psychological Association.

*Angus, R.G., Pearce, D.G., Buguet, A.G.C., & Olsen, L. (1979). Vigilance performance of men sleeping under arctic conditions. *Aviation, Space, and Environmental Medicine, 50*(7), 692-696.

Aschoff, J. (1984). Circadian timing. In: J. Gibbon and L. Allan (Eds.), *Timing and time perception.* (pp. 442-468). New York, NY: The New York Academy of Sciences.

*Bateman, R.P. (1980). *The effects of increased ambient temperature on dual task performance*. Unpublished Doctoral Dissertation, Texas A&M University, College Station, Texas.

*Bell, C.R., & Provins, K.A. (1963). Relation between physiological responses to environmental heat and time judgments. *Journal of Experimental Psychology, 66(*6), 572-579.

*Bell, C.R. (1964). Climate, body temperature, and vigilance performance. *Proceeding of the Second International Congress on Ergonomics*, (pp. 169-172). London: Taylor & Francis.

*Bell, P.A. (1978). Effects of noise and heat stress on primary and subsidiary task performance. *Human Factors, 20*(6), 749-752.

*Bell, P.A. (1982). Effects of heat, social facilitation, sex differences, and task difficulty on reaction time. *Human Factors, 24*(1), 19-24.

*Beshir, M.Y., El-Sabagh, A.S., & El-Nawawi, M.A. (1981). Time on task effect on tracking performance under heat stress. *Ergonomics, 24*(2), 95-102.

Blagden, C. (1775a). Experiments and observations in an heated room. *Philosophical Transactions*, *65*, 111-123.

Blagden, C. (1775b). Further experiments and observations in an heated room. *Philosophical Transactions*, *65,* 484-494.

*Blockley, W.V., & Lyman, J.H. (1950). *Studies of human tolerance for extreme heat: III: Mental performance under heat stress as indicated by addition and number checking tests*. USAF Technical Report 6022, Ohio: Wright Paterson AFB.

Blockley, W.V., & Lyman, J.H. (1951). *Studies of human tolerance for extreme heat: IV: Psychomotor performance of pilots as indicated by a task simulating aircraft instrument flight*. Technical Report 6521, Ohio: Wright Paterson AFB.

Boregowda, S.C. (1999). Thermodynamic modeling and analysis of human stress responses. *Dissertation Abstracts International: Section B: The Sciences & Engineering,* 60(3-B), 1185.

Brief, R.S., & Confer, R.G. (1971). Comparison of heat stress indices. *American Industrial Hygiene Association Journal, 32,* 11-16.

*Bursill, A.E. (1958). The restriction of peripheral vision during exposure to hot and humid conditions. *Quarterly Journal of Experimental Psychology, 10,* 113-129.

Burton, A.C., & Edholm, O.G. (1955). *Man in a cold environment: Physiological and pathologic effects of exposure to low temperatures*. London: Edward Arnold.

Cannon, W.B. (1932). *The wisdom of the body*. New York: W.W. Norton.

Cheung, S.F., & Chan, D.K.S. (2004). Dependent effect sizes in meta-analysis: Incorporating the degree of interdependence. *Journal of Applied Psychology, 89*, 780-791.

*Chiles, W.D. (1958). Effects of elevated temperatures in performance of a complex mental task. *Ergonomics, 2*, 89-96.

*Cian, C., Barraud, P.A., Melin, B., & Raphel, C. (2001). Effects of fluid ingestion on cognitive function after heat stress or exercise-induced dehydration. *International Journal of Psychophysiology, 42*, 243-251.

*Cian, C., Koulmann, N., Barraud, P.A., Raphel, C., Jimenez, C., & Melin, B. (2000). Influences of variations in body hydration on cognitive function: Effect of hyperhydration, heat stress, and exercise-induced dehydration. *Journal of Psychophysiology, 14*, 29-36.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Lawrence Erlbaum Associates.

*Colquhoun, W.P., & Goldman, R.F. (1972). Vigilance under induced hypothermia. *Ergonomics, 15*(6), 621-632.

*Courtright, J.F. (1976). *Effects of whole and partial body exposure to dry heat on certain performance measures.* Unpublished Master's Thesis, Purdue University, West Lafayette, Indiana.

*Curley, M.D., & Hawkins, R.N. (1983). Cognitive performance during a heat acclimatization regimen. *Aviation, Space, and Environmental Medicine, 54*(8), 709-713.

Ellis, H. (1758). An account of the heat of the weather in Georgia. *Philosophical Transactions*, *50*, 754-756.

*Ellis, H.D. (1982). The effects of cold on the performance of serial choice reaction time and various discrete tasks. *Human Factors, 24*(5), 589-598.

*Ellis, H.D., Wilcock, S.E., & Zaman, S.A. (1985). Cold and performance: The effects of information load, analgesics, and the rate of cooling. *Aviation, Space, and Environmental Medicine, 56*, 233-237.

Enander, A. (1984). Performance and sensory aspects of work in cold environments: A review. *Ergonomics, 27*(4), 365-378.

Enander, A. (1986). *Sensory reactions and performance in moderate cold*. Unpublished Doctoral Dissertation, University of Uppsala, Uppsala, Sweden.

*Enander, A. (1987). Effects of moderate cold on performance of psychomotor and cognitive tasks. *Ergonomics, 30*(10), 1431-1445.

Enander, A.E. (1989). Effects of thermal stress on human performance. *Scandinavian Journal of Work, Environment, and Health, 15*(supp1), 27-33.

Enander, A.E., & Hygge, S. (1990). Thermal stress and human performance. *Scandinavian Journal of Work, Environment, and Health, 16*(supp1), 44-50.

*Epstein, Y., Keren, G., Moisseiev, J., Gasko, O., & Yachin, S. (1980). Psychomotor deterioration during exposure to heat. *Aviation, Space, and Environmental Medicine, 51*(6), 607-610.

*Faerevik, H., & Reinertsen, R.E. (2003). Effects of wearing aircrew protective clothing on physiological and cognitive responses under various ambient conditions. *Ergonomics, 46*(8), 780-799.

Fanger, P.O. (1967). *Thermal comfort*. Copenhagen: Danish Technical Press.

Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381-391.

Gaillard, A.W.K. (2005, July). *Concentration — An instrument to augment cognition*. Paper presented at the 1st International Conference on Augmented Cognition, Las Vegas, NV.

*Givoni, B. & Rim, Y. (1962). Effect of the thermal environment and psychological factors upon subjects' responses and performance of mental work. *Ergonomics, 5*, 99-114.

Goldman, R.F. (2001). Introduction to heat-related problems in military operations. In: K.B. Pandolf, R.E. Burr, C.B. Wenger, and R.S. Pozos (Eds.), *Medical aspects of harsh environments.* (pp. 3-49). Washington D.C.: Borden Institute.

Goodman, D., Hancock, P.A., Runnings, D.W., & Brown, S.L. (1984). Temperature induced changes in neuromuscular function: Central and peripheral mechanisms. *Perceptual & Motor Skills*, *59*, 647-656.

Grether, W.F. (1973). Human performance at elevated environmental temperatures. *Aerospace Medicine, 44*, 747-755.

*Griffiths, I.D., & Boyce, P.R. (1971). Performance and thermal comfort. *Ergonomics, 14*(4), 457-468.

Hancock, P.A. (1980). Simulated and experimental temperature responses in man during exercise in varying environments. *Computers in Biology and Medicine*, *10*, 1-9.

Hancock, P.A. (1981a). The simulation of human core temperature. *International Journal of Bio-Medical Computing*, *12,* 59-66.

Hancock, P.A. (1981b). Heat stress impairment of mental performance: A revision of tolerance limits. *Aviation*, *Space and Environmental Medicine*, *52,* 177-180.

Hancock, P.A. (1982). Task categorization and the limits of human performance in extreme heat. *Aviation*, *Space and Environmental Medicine*, *53,* 778-784.

Hancock, P.A. (1984). Effect of environmental temperature on display monitoring performance: An overview with practical implications. *American Industrial Hygiene Association Journal*, *45*, 122-126.

Hancock, P.A. (1986a). Sustained attention under thermal stress. *Psychological Bulletin*, *99*, 263-281.

Hancock, P.A. (1986b). The effect of skill on performance under an environmental stressor. *Aviation*, *Space*, *and Environmental Medicine*, *57*, 59-64.

Hancock, P.A. (2003). The ergonomics of torture: The moral dimension of evolving human-machine technology. *Proceedings of the Human Factors and Ergonomic Society*, 47, 1009-1011.

Hancock, P.A. (2006). Thermal comfort. In: W. Karwowski (Ed.). *International Encyclopedia of Ergonomics and Human Factors.* (Second Edition). (pp. 1859-1862), London: Taylor & Francis.

Hancock, P.A., & Desmond, P.A. (2001) (Eds.). *Stress, workload and fatigue.* Mahwah, NJ: Lawrence Erlbaum.

Hancock, P.A., & Ganey, H.C.N. (2003). From the inverted-U to the extended-U: The evolution of a law of psychology. *Human Performance in Extreme Environments*, *7* (1), 5-14.

Hancock, P.A., & Pierce, J.O. (1985). Combined effects of heat and noise on human performance: A review. *American Industrial Hygiene Association Journal*, *46*, 555-566.

Hancock, P.A., & Vasmatzidis, I. (1998). Human occupational and performance limits under stress: The thermal environment as a prototypical example. *Ergonomics*, *41*, 1169-1191.

Hancock. P.A., & Vasmatzidis, I. (2003). Effects of heat stress on cognitive performance: The current state of knowledge. *International Journal of Hyperthermia*, *19*, 355-372.

Hancock, P.A., & Vercruyssen, M. (1988). Limits of behavioral efficiency for workers in heat stress. *International Journal of Industrial Ergonomics*, *3*, 149-158.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519-537.

Harris, W.C. Hancock, P.A., & Harris, S.C. (2005). Information processing changes following extended stress. *Military Psychology*, *17(*2), 115-128.

Hebb, D.O. (1955). Drives and the C.N.S. (Conceptual Nervous System). *Psychological Review*, 62, 243-254.

Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Hedges, L.V., Shymansky, J.A., & Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis.* Washington, DC: National Science Teachers Association.

Hockey, G.R.J., Gaillard, A.W.K., & Coles, M.G.H. (1986). (Eds.). *Energetics and human information processing*. Martinus Nijhoff Publishers: Dordrecht, Germany.

*Hocking, C., Silberstein, R.B., Lau, W.M., Stough, C., & Roberts, W. (2001). Evaluation of cognitive performance in the heat by functional brain imaging and psychometric testing. *Comparative Biochemistry and Physiology, 128*(A), 719-734.

Hoff, H.H., Guillemin, L., & Guillemin, R. (1967). *The cahier rouge of Claude Bernard*. Cambridge, MA: Schenkman Publishing Co.

*Holmberg, I., & Wyon, D. (1969). The dependence of performance in school on classroom temperature. *Educational and Psychological Interactions, 31*, 1-20.

Huffcutt, A.I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology, 80*(2), 327-334.

Hunter, J.E., & Schmidt, F.L. (2004). *Methods of meta-analysis: Correcting for error and bias in research findings. 2nd Ed.* Thousand Oaks, CA: Sage.

Huntington, E. (1919). *World-power and evolution*. New Haven, CT: Yale University Press.

*Hygge, S., & Knez, I. (2001). Effects of noise, heat and indoor lighting on cognitive performance and self-reported affect. *Journal of Environmental Psychology, 21*, 291-299.

Johnson, R.F., & Kobrick, J.L. (2001). Psychological aspects of military performance in hot environments. In: K.B. Pandolf, R.E. Burr, C.B. Wenger, and R.S. Pozos (Eds.), *Medical aspects of harsh environments.* (pp. 135-159). Washington D.C.: Borden Institute.

Kleitman, N. (1939/1963). *Sleep and wakefulness*. Chicago University Press: Chicago.

Lachman, R., Lachman, J.L., & Butterfield, E.C. (1979). *Cognitive psychology and information processing*. Hillsdale, NJ: Erlbaum.

Lazarus, R.S., & Folkman, S. (1984). *Stress, appraisal and coping*. New York, USA: Springer Publishing Company.

Lind, A.R. (1963). A physiological criterion for setting thermal environmental limits on everyday work. *Journal of Applied Physiology, 18*, 51-56.

Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Lockhart, J.M. (1968). Extreme body cooling and psychomotor performance. *Ergonomics, 11*(3), 249-260.

*Lockhart, J.M., & Kiess, H.O. (1971). Auxiliary heating of the hands during cold exposure and manual performance. *Human Factors, 13*(5), 457-465.

Loeb, M. & Jeantheau, G. (1958). The influence of noxious environmental stimuli on vigilance. *Journal of Applied Psychology, 42*, 47-49.

*Mackworth, N.H. (1946). Effects of heat on wireless telegraphy operators hearing and recording morse messages. *British Journal of Industrial Medicine, 3*, 143-158.

Martinussen, M., & Bjornstad, J.F. (1999). Meta-analysis calculations based on independent and non-independent cases. *Educational and Psychological Measurement, 59*, 928-950.

*Mayo, G.D. (1955). Effect of temperature upon technical training. *Journal of Applied Psychology, 39*(4), 244-246.

*Moreland, S., & Barnes, J.A. (1970). *Exploratory study of pilot performance during high ambient temperatures/humidity*. Aberdeen Research and Development Technical Report 5184.12.646, Maryland: Fort Aberdeen.

Morris, S.B., & DeShon, R.P. (2002). Combining effect size estimate in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105-125.

Näätänen, R. (1973). The inverted-U relationship between activation and performance. In: S. Kornblum (Ed.). *Attention and performance IV*. (pp. 155-174), Handan: Academic Press.

Parsons, K.C. (1993). *Human thermal environments: The effects of hot, moderate and cold environments on human health, comfort and performance*. London, England: Taylor & Francis.

Parsons, K.C. (1995). International heat stress standards: A review. *Ergonomics, 38,* 6-22.

Patterson, M.J., Taylor, N.A.S., & Amos, D. (1998). *Physical work and cognitive function during acute heat exposure before and after heat acclimization* (DSTO-TR-0683). Melbourne, Australia: DSTO Aeronatuical and Maritime Research Laboratory.

*Peccolo, C. (1962). *The effect of thermal environment on learning: A pilot study*. Unpublished Doctoral Dissertation, University of Iowa. Iowa City, IA.

*Pepler, R.D. (1960). Warmth, glare and a background of quiet speech: A comparison of their effects on performance. *Ergonomics, 3*, 68-73.

*Pepler, R.D., & Warner, R.E. (1968). Temperature and learning: An experimental study. *Proceeding of the ASHRAE, 74*, 211-219.

Pilcher, J., Nadler, E., & Busch, C. (2002). Effects of hot and cold temperature exposure on performance: A meta-analytic review. *Ergonomics, 45*(10), 682-698.

Poulton, E.C. (1965). On increasing the sensitivity of measures of performance. *Ergonomics*, *8*(1), 69-76.

Poulton, E.C. (1970). *Environment and human efficiency*. Springfield, IL: Charles C. Thomas.

Poulton, E.C. (1976). Arousing environmental stresses can improve performance, whatever people say. *Aviation, Space, & Environmental Medicine, 47*, 1193-1204.

*Poulton, E.C., & Kerslake, D.M. (1965). Initial stimulating effect of warmth upon perceptual efficiency. *Aerospace Medicine, 36*, 29-32.

Prosser, C.L., & Nelson, D.O. (1981). The role of nervous systems in temperature adaptation of poikilotherms. *Annual Review of Physiology*, *43,* 281-300.

Ramsey, J.D. (1995). Task performance in heat: A review. *Ergonomics, 38*(1), 154-165.

Ramsey, J.D. & Morrissey, S.J. (1978). Isodecrement curves for task performance in hot environments. *Applied Ergonomics*, *9*, 66-72.

*Razmjou, S. (1992). Sustained attention and serial responding in heat: Mental effort in the control of performance. *Aviation, Space, and Environmental Medicine, 63*, 594-601.

*Razmjou, S. (1996). Mental workload in heat: Toward a framework for analyses of stress states. *Aviation, Space, and Environmental Medicine, 67*(6), 530-538.

*Reading, J.E., Kincaid, P.S., Roberts, D.E., Hesslink, R.L., & Pozos, R.S. (1994). *The effect of shivering on rifle shooting performance in U.S. Marines*. US Naval Health Research Center Technical Report 94-5, California: San Diego.

*Reardon, M.J., Fraser, E.B., & Omer, J.M. (1998). Flight performance effects of thermal stress and two aviator uniforms in UH-60 helicopter simulators. *Aviation, Space, and Environmental Medicine, 69*(6), 569-576.

*Reddy, S.P. (1974). *Sedentary job performance within the thermal comfort zone*. Unpublished Master's Thesis, Texas Tech University, Lubbock, Texas.

*Reilly, R.E., & Parker, J.F. (1988). *Effect of heat stress and prolonged activity on perceptual-motor performance*. NASA Technical Report CR-1153, Virginia: Arlington.

Rhudy, J.L., & Meagher, M.W. (2003). Negative affect: Effects on an evaluative measure of human pain. *Pain*, *104*, 617-626.

Seyle, H. (1976). *The stress of life.* New York: McGraw-Hill.

*Sharma, V.M., & Panwar, M.R. (1987). Variations in mental performance under moderate cold stress. *International Journal Biometeorology, 31*(1), 85-91.

Southern Region Headquarters National Weather Service. Relative humidity and dewpoint temperature from temperature and wet-bulb temperature. Retrieved May 15, 2005, from http://www.srh.noaa.gov/elp/wxcalc/formulas/rhTdFromWetBulb.html

*Tanaka, M., Tochihara, Y., Yamazaki, S., Ohnaka, T., & Yoshida, K. (1983). Thermal reaction and manual performance during cold exposure while wearing cold-protective clothing. *Ergonomics, 26*(2), 141-149.

Taylor, C.L. (1948). *Interim Report*, Committee on Aviation Medicine. Washington, D.C.: National Research Council.

*Teichner, W.H., & Wehrkamp, R.F. (1954). Visual-motor performance as a function of short-duration ambient temperature. *Journal of Experimental Psychology, 47,* 447-450.

*Thomas, J.R., Ahlers, S.T., House, J.F., & Schrot, J. (1989). Repeated exposure to moderate cold impairs matching-to-sample performance. *Aviation, Space, and Environmental Medicine, 60*(11), 1063-1067.

*Tikuisis, P., Keefe, A.A., Keillor, J., Grant, S., & Johnson, R.F. (2002). Investigation of rifle marksmanship on simulated targets during thermal discomfort. *Aviation, Space, and Environmental Medicine, 73*(12), 1176-1183.

Ursin, H., & Eriksen, H.R. (2004). The cognitive activation theory of stress. *Psychoneuronendocrinology, 29*(5), 567-592.

*van Orden, K.F., Benoit, S.L., & Osga, G.A. (1996). Effects of cold air stress on the performance of a command and control task. *Human Factors, 38*(1), 130-141.

Vasmatzidis, I., Schlegel, R.E., & Hancock, P.A. (2002). Dual task performance under heat stress. *Ergonomics*, *45* (3), 218-239.

Vernon, H.M., & Bedford, T. (1930). *A study of heating and ventilation in schools*. Industrial Health Research Board, Medical Research Council Report, HMSO: London, England.

Von Humboldt, A. (1816). *Voyage aux regions equinoxiales*. Paris: Greek, Latin, & German, Library.

*Weiner, J.S., & Hutchinson, J.C.D. (1945). Hot humid environment: Its effect on the performance of a motor co-ordination test. *British Journal of Industrial Medicine*, *2*, 154-157.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering psychology and human performance*. Upper Saddle River, NJ: Prentice Hall.

Wyndham, C.H. (1969). Adaptation to heat and cold. *Environmental Research*, *2*, 442-469.

Wyndham, C.H., Strydom, N.B., Munro, A., MacPherson, R.K., Metz, B., Schaff, G., & Schieber, J. (1964). Heat reactions of Caucasians in temperate, in hot, dry, and in hot, humid climates. *Journal of Applied Physiology*, *19*, 607-612.

*Wyon, D.P. (1969). *The effects of moderate heat stress on the mental performance of children*. Stockholm, Sweden: National Swedish Institute for Building Research Technical Report D8/69.

Yaglou, C.P., & Minard, D. (1957). Control of heat casualties at military training centers. *American Medical Association Archives of Industrial Health, 16*, 302-316 and 405 (corrections).

Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482.

ACKNOWLEDGEMENT

The views expressed here are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government. This research was specifically supported by a Grant from the Army Research Laboratory through Micro Analysis and Design, P.A. Hancock and J.L. Szalma, *Principal Investigators* (Grant #64018042). We wish to thank Dr. Laurel Allender, John Lockett, Sue Archer, and Celine Richer, for providing administrative and technical direction for the Grant. This work was also facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program, P.A. Hancock, *Principal Investigator*, administered by the Army Research Office under Grant DAAD19-01-1-0621. The authors also wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the latter Grant. Further, we are very appreciative of the insightful comments of the three anonymous reviewers provided on an earlier version of this paper, for which we are indebted.

Table 1. Effects of thermal stress on performance divided into respective categories of analysis. Here, k represents the number of studies that fall into each respective category of analysis. $\overline{g}$ represents the mean effect size of that category. s_e^2 represents the sampling error variance. s_δ^2 represents the variance of the effect sizes. s_δ^2 represents the difference between s_δ^2 and s_e^2, and represents the variance in the effect sizes not due to sampling error (cf., Hunter & Schmidt, 2004).

Category	k	$\overline{g}$	s_e^2	s_g^2	s_δ^2	95% CI (s_e^2)
Global	56	-.34	.07	1.01	.94	-.41<δ<-.27
Heat	40	-.29	.06	.47	.41	-.36<δ<-.21
Cold	21	-.26	.09	2.57	2.48	-.39<δ<-.14
Heat ET>85 °F (29.4°C)	31	-0.25	0.08	0.67	.59	-0.35<δ<-0.16
Heat ET<85 °F (29.4°C)	14	-0.27	0.04	0.24	.20	-0.37<δ<-0.17
Below 78.3°F (25.7°C) ET	5	-.97	.13	.65	.52	-1.28<δ<-.65
78.3°F<85°F (25.7°C <29.4°C) ET	12	-.18	.02	.11	.09	-.26<δ<-.09
85°F<95.4°F (29.4°C <35.2°C) ET	14	-.40	.08	.21	.13	-.55<δ<-.26
Above 95.4°F (35.2°C) ET	21	-.39	.10	1.82	1.72	-.52<δ<-.25
Below 52°F (11.1°C) ET	12	-.78	.15	2.30	2.15	-.99<δ<-.35
Above 52°F (11.1°C) ET	9	.35	.05	1.27	1.22	.20<δ<.49
Perception	24	-.92	.14	1.82	1.68	-1.07<δ<-.77
Cognitive	30	-.18	.04	.51	.47	-.25<δ<-.11
Psychomotor	19	-.46	.10	2.01	1.91	-.60<δ<-.31
Accuracy	50	-.28	.07	.80	.73	-.35<δ<-.20
RT	28	-.45	.09	2.57	2.48	-.56<δ<-.34
Less than 1 hour	13	-.13	.04	1.01	.97	-.24<δ<-.02
1 to 2 hours	19	-.60	.14	2.34	2.20	-.77<δ<-.44
2 to 3 hours	17	-.76	.11	3.02	2.91	-.91<δ<-.60
More than 3 hours	8	-.15	.04	1.28	1.24	-.30<δ<-.01

Table 2. Listing of studies reflecting the four heat stress temperature ranges. Note that some papers contain multiple studies that were analyzed in this meta-analysis.

Report	Temp Range	Temperature	Duration (in hours)	Adjusted ES
Wyon, 1969	Below 78°F (25.6°C)	73.3°F (22.9°C)	2	.06
Pepler & Warner, 1968	Below 78°F (25.6°C)	77.4°F (25.2°C)	3	.16
Holmberg & Wyon, 1969 Experiment 1	Below 78°F (25.6°C)	75.5°F (24.1°C)	0.67	-.16
Allen & Fischer, 1978 Experiment 1	Below 78°F (25.6°C)	74°F (23.3°C)	0.67	-1.71
Bateman, 1980	Below 78°F (25.6°C)	73.7°F (23.2°C)	missing	-1.45
Bell, 1978	Between 78-85°F (25.6-29.4°C)	81.4°F (27.4°C)	missing	-.48
Wyon, 1969	Between 78-85°F (25.6-29.4°C)	79.2°F (26.2°C)	2	.46
Pepler & Warner, 1968	Between 78-85°F (25.6-29.4°C)	83.2°F (28.4°C)	3	.12
Hygge & Knez, 2001	Between 78-85°F (25.6-29.4°C)	79.2°F (26.2°C)	2	-1.16
Givoni & Rim, 1962	Between 78-85°F (25.6-29.4°C)	82.7°F (28.2°C)	2	.14
Mackworth, 1946	Between 78-85°F (25.6-29.4°C)	84.9°F (29.4°C)	3	.03
Mayo, 1955	Between 78-85°F (25.6-29.4°C)	83.7°F (28.7°C)	80	-.04
Holmberg & Wyon, 1969 Experiment 1	Between 78-85°F (25.6-29.4°C)	80.2°F (26.8°C)	0. 67	-.12
Holmberg & Wyon, 1969 Experiment 2	Between 78-85°F (25.6-29.4°C)	80.2°°F (26.8°C)	0. 67	-.55
Reddy, 1974	Between 78-85°F (25.6-29.4°C)	81.2°F (27.4°C)	2	-.09
Moreland & Barnes, 1970	Between 78-85°F (25.6-29.4°C)	85°F (29.4°C)	2	-.16
Teichner & Wehrkamp, 1954	Between 78-85°F (25.6-29.4°C)	83.6F (28.7°C)	missing	.08
Tikuisis, Keefe, Keillor, Grant, & Johnson, 2002	Between 85-95°F (29.4-35°C)	94.6°F (34.8°C)	2	.17
Bell, 1978	Between 85-95°F (29.4-35°C)	92.6°F (33.7°C)	missing	-.72
Pepler & Warner, 1968	Between 85-95°F (29.4-35°C)	89.3°F (31.9°C)	3	.11
Razmjou, 1992	Between 85-95°F (29.4-35°C)	94.9°F (35°C)	1. 33	-1.17
Bell, Loomis, & Cervone, 1982	Between 85-95°F (29.4-35°C)	95.9°F (35.5°C)	2.58	-.41
Faerevik & Reinertsen, 2003	Between 85-95°F	89.8°F	3	-.15

	(29.4-35°C)	(32.1°C)		
Givoni & Rim, 1962	Between 85-95°F (29.4-35°C)	92.9°F (33.9°C)	2	-.14
Weiner & Hutchinson, 1945	Between 85-95°F (29.4-35°C)	95.1°F (35.1°C)	missing	-.60
Mackworth, 1946	Between 85-95°F (29.4-35°C)	93.1°F (34°C)	3	-.20
Bateman, 1980	Between 85-95°F (29.4-35°C)	92.6°F (33.7°C)	2	-.20
Reilly & Parker, 1988	Between 85-95°F (29.4-35°C)	88.6°F (31.4°C)	6	.16
Courtright, 1976	Between 85-95°F (29.4-35°C)	99.5°F (37.5°C)	1.1	-1.77
Poulton & Kerslake, 1965	Between 85-95°F (29.4-35°C)	88.6°F (31.4°C)	0. 33	-.27
Epstein, Keren, Moisseiev, Gasko, & Yachin, 1980	Between 85-95°F (29.4-35°C)	99°F (37.2°C)	2.25	-.70
Cian, Barraud, Melin, & Raphel, 2001	Above 95°F (35°C)	124.7°F (51.5°C)	2	-.59
Cian, Doulmann, Barraud, Raphel, Jimenez, & Melin, 2000	Above 95°F (35°C)	124.7°F (51.5°C)	2	.04
Blockley & Lyman, 1950	Above 95°F (35°C)	185.6°F (85.4°C)	0.6, 0.57, 1.23	2.73
Razmjou, 1996	Above 95°F (35°C)	104.9°F (40.5°C)	1.3	-.05
Curley & Hawkins, 1983	Above 95°F (35°C)	100°F (37.8°C)	2.58	1.10
Hocking, Silberstein, Lau, Stough, & Roberts, 2001	Above 95°F (35°C)	100.6°F (38.1°C)	missing	-1.94
Givoni & Rim, 1962	Above 95°F (35°C)	104°F (40°C)	2	-.04
Pepler, 1960 Experiment 2	Above 95°F (35°C)	108.4°F (42.4°C)	0.67	-1.18
Mackworth, 1946	Above 95°F (35°C)	102°F (38.9°C)	3	-3.94
Bell & Provins, 1963 Experiment 2	Above 95°F (35°C)	117.F (47.2°C)	2.75	.72
Bateman, 1980	Above 95°F (35°C)	103°F (39.5°C)	2	-.57
Bell, 1964 Experiment 1	Above 95°F (35°C)	140°F (60°C)	missing	.04
Bell, 1964 Experiment 2	Above 95°F (35°C)	132.8°F (56°C)	missing	.04
Chiles, 1958 Experiment 1	Above 95°F (35°C)	101°F (38.4°C)	1	-.01
Chiles, 1968 Experiment 2	Above 95°F (35°C)	109.1°F (42.8°C)	1	.16
Colquhoun & Goldman, 1972	Above 95°F (35°C)	113.9°F (45.5°C)	2	.00

Bursill, 1958 Experiment 1	Above 95°F (35°C)	120.1°F (48.9°C)	1.73	-.47
Bursill, 1958 Experiment 2	Above 95°F (35°C)	120.1°F (48.9°C)	1.73	-.88
Epstein, Keren, Moisseiev, Gasko, & Yachin, 1980	Above 95°F (35°C)	124.9°F (51.6°C)	2.25	-2.83
Reardon, Fraser, & Omer, 1998	Above 95°F (35°C)	100.8°F (38.2°C)	1.78	-.69
Teichner & Wehrkamp, 1954	Above 95°F (35°C)	100.7°F (38.2°C)	missing	.61

Table 3. Table of reports with heat acclimatization and core body temperatures.

Paper	Heat Acclimization?	If yes, how long?
Allen & Fisher, 1978 (exp. 1)	No	
Bateman, 1980	No	
Bell & Provins, 1963	No	
Bell, 1964 (exp. 1)	No	
Bell, 1964 (exp. 2)	No	
Bell, 1978	No	
Bell, Loomis, & Cervone, 1982	No	
Blockley & Lyman, 1950	No	
Bursill, 1958 (exp. 1)	Yes	Artificially heat-acclimatized over a period of a fortnight by exposing them for five days in one week to a condition of 95/85 F and a subsequent five days to 105/95 F for three hours daily, with an air velocity of 120 ft/min.
Bursill, 1958 (exp. 2)	Yes	Artificially heat-acclimatized over a period of a fortnight by exposing them for five days in one week to a condition of 95/85 F and a subsequent five days to 105/95 F for three hours daily, with an air velocity of 120 ft/min.
Chiles, 1958 (exp. 1)	No	
Chiles, 1958 (exp. 2)	No	
Cian, Barraud, Melin, & Raphel, 2001	No	
Cian, Koulmann, Barraud, Raphel, Jimenez, & Melin, 2000	No	
Colquhoun & Goldman, 1972	No	
Courtright, 1976	No	
Curley & Hawkins, 1983	Yes	10-day acclimatization
Epstein, Karen, Moisseiev, Gasko, & Yachin, 1980	No	
Faerevik & Reinertsen, 2003	No	
Givoni & Rim, 1962	No	
Hockey, Silberstein, Lau, Stough, & Roberts, 2001	No	
Holmberg & Wyon, 1969 (exp. 1)	No	
Holmberg & Wyon, 1969 (exp. 2)	No	
Hygge & Knez, 2001	No	
Mackworth, 1946	No	
Mayo, 1955	Yes	7 month course in electronics, two classes held at different temperatures and compared for end of year grades
Moreland & Barnes, 1970	No	
Peccolo, 1962	Yes	Experimental conditions those standard in the classroom, that children are already use to (i.e., acclimatized to).

Pepler & Warner, 1968	No	
Pepler, 1960 (exp. 2)	No	
Poulton & Kerslake, 1965	No	
Razmjou, 1992	No	
Razmjou, 1996 (exp. 2)	No	
Reardon, Fraser, & Omer, 1998	Unclear	Unclear, "activities during the first week included uniform and helmet fitting, simulator training, and heat stress acclimatization in an environmental chamber."
Reddy, 1974	No	
Reilly & Parker, 1988	No	
Teichner & Wehrkamp, 1954	No	
Tikuisis, Keefe, Keillor, Grant, & Johnson, 2002	No	
Weiner & Hutchinson, 1945	Yes	"In the main series of experiments the 6 subjects, aged 25 to 35, were well acclimatized to working in hot humid environments."
Wyon, 1969	No	

Table 4. Duration moderator analysis after removal of outliers/influential data points

Category	k	$\overline{g}$	s_e^2	s_g^2	s_δ^2	95% CI (s_e^2)
Less than 1 hour	11	-.28	.09	.83	.74	-.45<δ<-.10
1 to 2 hours	17	-.71	.14	.81	.67	-.89<δ<-.53
2 to 3 hours	13	-.41	.09	.45	.36	-.57<δ<-.25
More than 3 hours	5	-.57	.10	.30	.20	-.85<δ<-.29

Table 5. Effect of thermal stress as a function of task category and performance measure.

Category	k	$\overline{g}$	s_e^2	s_g^2	s_δ^2	95% CI (s_e^2)
HEAT						
Perception	17	-.78	.12	2.04	1.92	-.95<δ<-.62
Cognitive	20	-.23	.04	.44	.40	-.31<δ<-.15
Psychomotor	14	-.31	.09	.40	.32	-.47<δ<-.15
Accuracy	36	-.33	.06	.58	.52	-.41<δ<-.25
RT	20	-.26	.08	1.79	1.71	-.38<δ<-.13
COLD						
Perception	8	-1.13	.17	.76	.59	-1.42<δ<-.84
Cognitive	12	.41	.05	.98	.93	.28<δ<.54
Psychomotor	8	-.42	.08	4.08	4.00	-.62<δ<-.23
Accuracy	18	.05	.07	1.22	1.15	-.07<δ<.18
RT	12	-.11	.07	3.79	3.72	-.25<δ<.04

Table 6. Thermal stress papers resulting in positive effect sizes for cold in the cognitive task category.

Report	Adjusted Effect Size
Griffiths & Boyce, 1971	.03
Ellis, Wilcock, & Zaman, 1985	.82
Ellis, Wilcock, & Zaman, 1985	.21
Ellis, 1982	.26
Ellis, 1982	.81
Enander, 1987	.00
Pepler & Warner, 1968	1.33
Sharma & Panwar, 1987	1.84
Reddy, 1974	.16

Table 7. Thermal stress effects as functions of task category and temperature above and below the 85°F ET threshold of the 'prescriptive' zone, (Lind, 1963). Performance is differentiated by processing task demand and by changes in response capacity as expressed in measures of speed and accuracy.

Category	k	$\overline{g}$	s_e^2	s_g^2	s_δ^2	95% CI (s_e^2)
Perception						
Heat ET>85 °F (29.4°C)	17	-.78	.12	2.04	1.92	-.95<δ<-.62
Heat ET<85 °F (29.4°C)	---	---	---	---	---	---
Cognitive						
Heat ET>85 °F (29.4°C)	12	-.01	.05	.69	.64	-.13<δ<.12
Heat ET<85 °F (29.4°C)	12	-.34	.04	.30	.26	-.45<δ<-.23
Motor						
Heat ET>85 °F (29.4°C)	12	-.32	.09	.50	.41	-.49<δ<-.14
Heat ET<85 °F (29.4°C)	4	.01	.11	.05	---	-.20<δ<.22
Heat ET>85 °F (29.4°C)						
Accuracy	27	-.37	.09	.89	.80	-.49<δ<-.26
RT	16	-.33	.09	2.47	2.38	-.47<δ<-.18
Heat ET<85 °F (29.4°C)						
Accuracy	13	-.30	.04	.28	.24	-.40<δ<-.19
RT	7	.08	.02	.16	.14	-.03<δ<.18

Table 8. Time by temperature effects for heat and cold stress

Category	k	$\overline{g}$	s_e^2	s_g^2	s_δ^2	95% CI (s_e^2)
HEAT						
Less Than ET 85°F (29.4°C)						
< 120 min.	7	-.47	.05	.23	.18	-.64<δ<-.31
> 120 min.	3	.03	.01	.01	0	-.08<δ<.14
More Than ET 85°F (29.4°C)						
< 120 min.	19	-.29	.09	.67	.58	-.43<δ<-.15
> 120 min.	7	-.03	.05	.57	.52	-1.19<δ<.13
COLD						
Less Than ET 52°F (11.1°C)						
< 120 min.	12	-.71	.15	2.10	1.95	-.93<δ<-.50
> 120 min.	2	.13, -6.03*				
More Than ET 52°F (11.1°C)						
< 120 min.	4	.66	.11	1.61	1.50	.33<δ<.99
> 120 min.	4	.42	.02	1.46	1.44	.27<δ<.57

** Since only two studies exist in this category, each effect size is reported rather than the mean.*

Table 9. The breakdown of heat and cold stress effects by task category and performance measure

Category	k	$\overline{g}$	s_e^2	s_g^2	s_δ^2	95% CI (s_e^2)
Heat						
Perception						
Accuracy	12	-.41	.10	.27	.17	-.59<δ<-.22
RT	11	-.91	.13	3.13	3.00	-1.12<δ<-.70
Cognitive						
Accuracy	19	-.27	.04	.57	.53	-.36<δ<-.18
RT	4	.02	.03	.01		-.14<δ<.19
Psychomotor						
Accuracy	13	-.59	.13	.72	.59	-.78<δ<-.40
RT	5	.68	.13	1.32	1.19	.37<δ<1.00
Cold						
Perception						
Accuracy	6	-1.07	.16	1.20	1.04	-1.39<δ<-.74
RT	5	-.85	.14	.86	.72	-1.18<δ<-.52
Cognitive						
Accuracy	10	.05	.05	.56	.51	-.09<δ<.19
RT	6	.64	.19	.47	.28	.29<δ<.99
Psychomotor						
Accuracy	5	.58	.09	1.78	1.69	.32<δ<.85.
RT	4	-1.10	.44	3.07	2.63	-1.75<δ<-.45

BIOGRAPHIES

Peter A. Hancock is the Provost Distinguished Research Professor at the University of Central Florida, Orlando, FL. He received his Ph.D. in Motor Performance from the University of Illinois at Urbana-Champaign in 1983.

Jennifer M. Ross is a fourth year graduate student at the University of Central Florida in the Applied Experimental Human Factors Ph.D. Program. Jennifer earned her M.A. in Modeling and Simulation from the University of Central Florida.

James L. Szalma is a Research Scientist at the University of Central Florida. He received his Ph.D. in Applied Experimental Psychology/Human Factors from the University of Cincinnati in 1999.

Running heads (verso) *G.E. Conway et al.*
(recto) *Whole-body vibration effects on performance*

Article Type: Review Article (Meta-analysis)

A quantitative meta-analytic examination of whole-body vibration effects on human performance

G.E. CONWAY*, J.L. SZALMA and P.A. HANCOCK

[A1] MIT[2] Lab, University of Central Florida, 3100 Technology Parkway, Suite 337, Orlando, FL, 32826, USA.

*Corresponding author. Email: gconway@ist.ucf.edu

Whole-body vibration exerts a substantive influence in many work environments. The primary objective for this work was to quantify such effects by identifying those moderating variables that influence the degree to which performance is affected. To achieve this, a comprehensive meta-analysis was conducted which synthesised the existing research evidence. A total of 224 papers and reports were identified, and from these, 115 effect sizes were derived from 13 experiments which survived the screening procedure. Results indicate that vibration acts to degrade the majority of goal-related activities, especially those with high demands on visual perception and fine motor control. Gaps in the current research literature are identified and suggestions offered with regard to a more theoretically-driven approach to testing vibration effects on human performance.

Keywords: Whole-body vibration; Human performance; Stress, Meta-analysis

1 Introduction

The problem of vibration in work environments is an issue that has long been of concern (e.g. Ramazzini, 1713). Though vibration disturbance to performance may be thought of as an 'old' problem, its pervasive effects are still recognised by contemporary engineers and behavioural scientists as a key ergonomic issue (see Peacock, Chase, *et al*. 2005; Fraser, Svensson, *et al*. 2004). Vibration has traditionally been conceived in two distinct forms. The first of these forms concerns effects to single, exposed limbs. The second, but by far the more pervasive form concerns vibration of the whole body ('whole body vibration', WBV). It is the latter category which is the focus of this work. Clear examples of the potential disruptive influence of WBV can be seen in any number of work environments but perhaps most significantly in those that require the concomitant use of both *transportation* and *information* systems. Operator performance within aviation, maritime, and land-based vehicle operations are therefore constantly and consistently at risk from WBV effects (Mansfield, 2004).

Vibration takes the form of a mechanical wave which, in a manner analogous to noise effects, serves to transfer energy but not matter (Mansfield, 2004). It is this energy transfer that proves problematic to performance since it is transmitted to, and dissipated within the human that is in contact with the vibrating structure (Griffin, 2004). Therefore WBV manifests itself most often when vibrating surfaces, such as a vehicle in motion, support the human operator. Immediate effects only cease when the supporting structure itself stops vibrating, or when a decoupling between the user and the surface of support occurs. However, this immediate cessation of direct effects does not eliminate the possibility of after-effects which are common occurrences in many other forms of stress exposures (e.g. Cohen, 1980).

Vibration is a particular threat as it imposes a direct mechanical influence on several aspects of task response. If the dynamic adaptability model of stress and performance (Hancock & Warm, 1989) is used as a framework for understanding such effects, then it can be seen that vibration manifests itself in each of the three components that make up the identified 'trinity of stress' (see Figure 1). Firstly vibration is a function of the ambient environment at any given time, and so can affect input-related activities (such as the collection of information through the different sensory modalities – mainly vision). Vibration can also influence 'output' response processes, reducing the effectiveness and efficiency of motor performance. Vibration also exerts an influence on adaptive processes - the individual's attempts to cope with their present environmental demands. Though vibration is known to have a pervasive influence on performance, the magnitudes of these effects are known to change according to the impact of a number of moderating factors.

[Insert Figure 1 about here]

1.1 Whole-body vibration and performance - the influence of moderating factors

Two specific groups of moderators are the *characteristics of the vibration itself* and the *characteristics of the task* at hand. These are discussed in turn. Vibration occurs and is measured in three translational *axes* in accordance with the standard biodynamic coordinate system (International Organization for Standardization, 1997). Though recognition of these three basic axes of vibration is important, it is even more valuable to consider the interaction of vibration axis with other factors. For example, the effect of the axis of vibration on performance is often dependent on task characteristics (e.g. which directions are the most important in a 2-dimensional tracking task) or the frequency of the vibration (given that there may be differential performance effects of the three axes at different frequencies).

The second vibration characteristic to be considered is that of *waveform*. Vibration may be random, intermittent (non-periodic) or continuous (periodic). The most common vibration form encountered in the literature is that of continuous sinusoidal vibration, which consists of repeated cycles where the object or person oscillates about their original position. Within the sinusoidal wave, variation can occur in how far the object moves from its origin in the cycle (the *magnitude*), the number of complete cycles per second (the *frequency*), and in the amount of time for which the vibration persists (when this is in relation to the performer it is termed *exposure duration*).

The *magnitude* of vibration represents how far the vibrating object moves from its starting point at the extremes of each cycle. Though the magnitude of vibration transferred to the body is linearly related to the vibration from the supporting surface, the implications of this for task performance depends upon the interaction with the task characteristics and the vibration frequency (Griffin, 1992).

The *frequency* of vibration is defined by the number of complete cycles per second (Hz). Again, it is difficult to quantify the effects of frequency in isolation; of specific interest here are the interactions with the direction of the vibration and the body parts that are used during performance (either for sensory or response processes). The latter interaction is of particular concern when we understand that different parts of the human body each possess their own individual ‘resonance’ frequency. *Resonance* refers to frequencies of vibration at which the effects on a given body part are maximised (i.e. the effects of the vibration on the body part are higher than would be expected if the vibration in the transmitting surface were studied in isolation).

The last characteristic of the vibration stimulus to be considered here is *exposure duration*. A key issue to recognise is that vibration (or any other stressor) effects take place within a four-dimensional environment, where the fourth dimension is *time*. Although the direct effects of vibration are well recognised, the full picture concerning the interaction of exposure duration and task type has yet to be elucidated.

Thus an important additional moderating factor is the type of task to be performed. The principle here being that WBV can have differential effects on different types of task. Though there is no one universally accepted taxonomy for performance type, one method

of categorising tasks is through accepted information processing models (e.g. see Wickens & Hollands, 2000). This approach usually assigns tasks to processing stages that are most characteristic of the demands made. For example, one task may place the primary demand on sensory processes; another task may emphasise decision making, and yet again a third task may hinge on the accurate use of fine motor control.

1.2 Purpose of current work

The introduction of numerous moderating factors can generate confusion with respect to performance effects. Further, current understanding of the mechanisms underpinning WBV effects specifically on information processing and the moderating influence of exposure duration is particularly low (Griffin, 2004). Accordingly, the construction of a theory that is capable of explaining the role of each moderating factor and the mechanisms underpinning their effects is an important challenge. Indeed, the current lack of a single accepted theory is testament to the scale of this problem. Advances in understanding may be achieved via two general strategies. Firstly, Griffin (2004) has emphasised that though an explanation of the numerous mechanisms underpinning WBV effects is perhaps too ambitious, efforts that specify the relative importance of the various factors are highly valuable. Secondly, WBV research may benefit from considering methodological and theoretical developments in parallel areas of research. For instance, benefits may be gained from comparative analysis of progress made in the general field of stress and performance research over the past half-century (Hancock & Desmond, 2001; Hockey, 1983). The transfer of knowledge from areas of stress research can result in the identification of alternative ways through which to reconsider old problems. For example, an emphasis may be placed on new methods, or the assessment of a broader range of impinging variables (Hockey & Hamilton, 1983).

Our current work therefore focuses on the first strategy, and attempts to provide a comprehensive, quantitative analysis of all available empirical studies which assessed the influence of whole-body vibration on human performance. Our goal is to determine the relative importance of the various moderating factors, and to identify gaps in understanding that may be exploited by both short- and long-term research strategies. We have quantified these effects using formal meta-analytic procedures (see Hedges & Olkin, 1985; Hunt, 1997; Hunter & Schmidt, 2004). There are many advantages associated with conducting a meta-analysis for this purpose. Meta-analysis provides a quantitative synthesis of the literature, allows effects drawn from different methods and measures to be combined, controls for sampling error and the low power of individual experiments, while permitting an examination of potential moderating variables (Hunter & Schmidt, 2004; Lipsey, 2003; Lipsey & Wilson, 2001).

2 Method

2.1 Literature accumulation

To collect the relevant studies for the WBV meta-analysis, an exhaustive literature search was performed using the PsycINFO®, MEDLINE®, and the Dissertation Abstracts International Database. The following search term combinations were used as a primary keyword: 'whole body', 'vibration', 'performance', 'cognition', 'motor', and 'vigilance'. In addition we used a number of web-based search engines, e.g. Google® and their specialist derivates, e.g. Google Scholar® to seek further references not found in the initial formal scan. After a preliminary listing of articles was obtained, additional articles were collected by surveying the reference lists from those already available and by retrospectively examining article citations through Science Citation Index® (SCI). Following this initial, formal search procedure, subject matter experts (SME's) were consulted concerning any remaining, pertinent references that may not have been identified by this primary search process. All such articles cited by the SME's were then searched for additional references. The exhaustion of these dual processes composed the present listing of articles. This effort resulted in the identification of two-hundred and twenty-four articles, reports, dissertations, and theses. We are, of course, well aware of the documented efficiency of such search procedures (e.g. Sommer, 1987), as well as the perennial concern for 'file drawer' effects (Hunter & Schmidt, 2004). From these collected works, eleven papers were identified which met the following selection criteria.

2.2 Criteria for inclusion

Each selected study had to report an empirical examination of vibration stress in which the experimental manipulation employed an application of WBV. Next, the study had to address directly the issue of WBV effects on performance. For example, studies that sought to disturb sleep through the application of vibration and then assess the effect of sleep-loss on performance were not considered suitable for inclusion (e.g. Arnberg, Bennerhult *et al*. 1990). Then, each report was required to include a control group for comparison purposes. Most frequently, this took the form of one group (the 'control') receiving no vibration, whilst the second group (the 'experimental' or 'treatment' group) were exposed to vibration. If a within-participant design was used, then each individual had to perform in both the non-vibration and the vibration conditions. Subsequently, each study had to report at least one measure of performance (e.g. marksmanship, manual dexterity, response time, tracking error). Studies using physiological or subjective response alone were thus excluded (e.g. Meister, Brauer, *et al*. 1984). Finally, each study had to include sufficient information regarding performance results to determine effect size estimates. Founded upon these five selection criteria, thirteen primary studies were accepted for use in the meta-analysis. The rejection of numerous primary studies in a meta-analysis is a common occurrence and necessary to insure meaningful data when combining effect size estimates across studies. In the present analysis, the modal reason for exclusion of a study was the lack of performance variables (the papers referred to here generally examine WBV influences on comfort, e.g. Suzuki, 1998). We collected a total of two-hundred and twenty-four studies. The screening process resulted in the qualification of thirteen of these studies for inclusion. These emerged from a total of

eleven different sources (i.e. articles, dissertations, technical reports, etc.), and resulted in one-hundred and fifteen effect sizes.

2.3 Calculation of effect size

Effect sizes for this study were the standardised mean difference between the experimental and the control conditions, often referred to as Hedge's *g* (Hedges & Olkin, 1985, p. 78; see also Hedges, Shymansky, *et al*. 1989). Many researchers are more familiar with Cohen's *d* (Cohen, 1988), which is conceptually similar to Hedge's *g* but has different distributional properties (see Hunter & Schmidt, 2004). When means and standard deviations were available the effect size was calculated by using the term expressed in equation (1):

$$g = \frac{\left(\overline{X}_E - \overline{X}_C\right)}{s} \quad \text{----------} \quad (1)$$

Where $\overline{X}_E$ = mean of the experimental condition, $\overline{X}_C$ = mean of the control condition, and s = standard deviation (for the control condition in within-subject designs; pooled standard deviation for between-subject designs). When the means and/or standard deviations were unavailable, effect sizes were computed with inferential statistics (e.g. *t*-ratios) or sums of squares/mean squares (e.g. from ANOVA tables) using equations presented by Lipsey and Wilson (2001) and Hedges, Shymansky, *et al*. (1989)

The *g* scores were adjusted for statistical bias, which decreases the accuracy of the estimates, particularly in cases where sample sizes are small. This adjustment (see Hedges & Olkin, 1985, pp. 78-81) provides an unbiased effect size.

$$d_u = g * 1 - \left(\frac{3}{4N - 9}\right) \quad \text{----------} \quad (2)$$

For E(g) x=N-2. The need for this correction comes from the statistical bias associated with *g* as an estimate of the effect size parameter (δ). The expectation for *g* is given by E(g)=δ/J(N-2), where

$$J(x) = 1 - \left(\frac{3}{4x - 1}\right) \quad \text{----------} \quad (3)$$

Thus, the expected value of *g* is not the population parameter δ but is proportional to it (see Hedges & Olkin, 1985). In calculating the effect size the directional sign of the effect size was controlled to ensure that positive scores represented better performance in the experimental group than the control group, whereas a negative score indicated worse performance. Vibration is a form of stress where the zero value on the ratio scale lies within the general comfort range of human tolerance (see Hancock & Warm, 1989).

In addition to the adjustments noted above, effect sizes derived from studies using within-subjects designs were adjusted to a between-subjects metric using procedures described by Morris and DeShon (2002). This correction was necessary in order to account for differences in standard deviation units as a function of experimental design.

2.4 Estimation of variances

The variance associated with each effect size was calculated using equations provided by Morris and DeShon (2002). These were combined for estimates of sampling error variance (s_e^2) and the variance among the observed effect sizes (s_g^2) using procedures described by Hunter and Schmidt (2004). These estimates can be used to derive the variance due to differences in the population effect size.(s_δ^2). Thus,

$$s_\delta^2 = s_g^2 - s_e^2 \quad \text{----------} \quad (4)$$

A large s_δ^2 indicates that there is variability among the observed effect sizes that cannot be accounted for by sampling error, and that there are likely to be one or more variables moderating the magnitude of the effect in question (see Hunter & Schmidt, 2004, pg. 288). Note that if all of the variance in the effect sizes were accounted for by sampling error, then s_δ^2 =0. The 95% confidence intervals reported was computed using the standard deviation corresponding to s_e^2

3 Meta-analytic results

Our first result concerns the rejection rate of studies that failed to meet the selection criteria. The collective survey of 224 studies and reports generated 13 useable studies from eleven papers. While it is true that this value varies with the threshold set by the selection criteria, we believe that our present criteria are not overly stringent. Hence, the extant empirical literature regarding WBV effects on performance is sparser than was initially anticipated. However, this attrition rate is similar to those reported in other published meta-analyses (e.g. Driskell & Mullen, 2005). The results of our meta-analysis are presented in Table 1.

[Insert Table 1 about here]

According to guidelines proposed by Cohen (1988), an effect size magnitude of 0.2 represents a ‘small’ effect, a level of 0.5 a ‘medium’ effect, and a magnitude of 0.8 represents a ‘large’ effect. Our global analysis shows an effect size of -0.95, which in meta-analytic terms represents a large overall deleterious effect. Since it includes 13 total studies, this represents a solid conclusion as a number of other meta-analyses have been reported using a similar numbers of studies (e.g. Driskell & Mullen, 2005).

Though the global analysis shows WBV generally causes decrement in performance, it is important to recognise that this outcome may be moderated by other factors. An additional, major aim here was therefore to establish what these factors are and to

quantitatively assess their effects. Our first moderator variable - the characteristics of the task - was assessed by categorising it into four components, namely *perceptual processes, cognitive processes, continuous fine motor actions* and *discrete fine motor actions*. An example of perceptual tasks would be in vigilance/target detection tasks (e.g. Warm, 1984). Working memory and mathematical reasoning are both examples of cognitive tasks (e.g. Baddeley, 1986). A continuous fine motor control task such as required in tracking, and a discrete fine motor control task such as seen in switch activations represent the output component of this differentiation.

Table 1 shows a large effect of WBV on perceptual task performance (k= 4, *g*= -1.79), with all four contributing studies reporting effects showing considerable performance decrements under WBV. Table 1 also shows a moderate effect size of WBV on cognitive performance (k= 1, *g*= -0.52). However, this result should be interpreted with caution, as it is drawn from the one study that met the inclusion criteria for this category. A greater number of studies were found to meet the criteria for motor performance, with 11 effect sizes found. The effect of WBV on *continuous fine motor* performance (k= 8, *g*= -0.89) was found to be similar to the effect on *discrete fine motor* performance (k= 3, *g*= -0.84). Both categories therefore exhibited large degradation effects under WBV. The effects for continuous fine motor performance appear more stable than those of discrete fine motor performance as reflected in the variability results.

To further examine the moderating influence of the task at hand, two different reflections of *response* were investigated, with the data categorised according to whether the *accuracy* of the performance or the *speed* of the performance was recorded. This analysis showed a greater decrement of WBV on the accuracy of performance (k= 9, *g*= -1.11) compared with the speed of performance (k= 5, *g*= -0.38).

To evaluate the differential effects of the WBV characteristics itself, the separate influences of *vibration intensity* and *vibration frequency* were assessed. There is no fundamental performance theory through which to derive differentiation of frequency and intensity effects (though such a rationale exists for comfort and health; see ISO, 1997). Therefore, we used a median split technique to differentiate this factor into the categories of either ‘high’ or ‘low’. From this median split, vibration intensity under 0.2 RMSg was considered to be low, while intensity over that threshold was considered as high. As is clear from Table 1, there was a smaller effect for low as compared to high intensity (i.e. k= 7, *g*= -0.67, and k= 9, *g*= -0.82, respectively). Examination of the frequency component of WBV was also performed using a median split technique. Values greater than 5 Hz represent high frequency WBV, and less than 5 Hz represent low frequency WBV. Table 1 reveals a similar trend to that observed for intensity. Thus, high frequencies exerting a substantial large negative effect on performance (k= 5, *g*= -1.98), while low frequencies exerted a moderate negative effect (k= 10, *g*= -0.65).

Another key factor in distilling WBV effects is *exposure duration*. The present results were categorised in two groups, studies that exposed performers to less than 30 min WBV and studies exceeding this duration. Shorter exposure times (< 30 min) produced a large effect on performance levels (k= 10, *g*= -0.87) while longer durations (> 30 min)

produced a substantively larger degradation (k= 3, *g*= -2.6). The latter effect however must again be interpreted with caution due to the large variation found in the individual effects contributing to this calculated effect size (s_e^2 = 9.91).

A secondary analysis was carried out to determine the extent to which exposure duration moderates the intensity effects of WBV on performance. Due to the very limited number of studies and the resultant low stability, the findings are again tentative but suggestive. Performance decrement did appear to increase as a function of both exposure duration and intensity. When short exposure durations were assessed as a function of low and high intensities, small (k= 5, *g*= -0.4) and then moderate (k= 8, *g*= -0.8) effects were found respectively. When the long exposure durations were assessed, a substantially large effect was found for long duration/low intensity combination (k= 2, *g*= -2.34), and an even larger effect was found for the combination of long duration/high intensity (k= 1, *g*= -3.84). Performance decrements therefore do appear to increase as duration and intensity increase in combination. This observation must, of course, be tempered by the fact that by the time we reach the highest levels of these combined factors, we are relying on the result of a single study (Loeb, 1954). While this does represent a concern, it is the present state of knowledge.

[Insert Table 2 about here]

4 Discussion

Our meta-analysis shows that WBV has a negative influence on performance. The degree to which the vibration was found to be disruptive was moderated by the type of task being performed. The largest effect was found for perceptual tasks (d= -1.79). It is unsurprising to find that perception is disrupted by a physical stressor such as vibration, which exerts its influence through motion effects on the human body. Further support for this contention is evident in the effect sizes for the continuous and discrete fine motor tasks (*g*= -0.89 and *g*= -0.84 respectively). Again, the general mechanism of performance disruption is readily apparent since these tasks require control of motor responses, which are easily disturbed when the vibration is absorbed within the body (Griffin, 2004). Though a moderate negative effect was found for WBV on cognitive performance (*g*= -0.52), the effect size was derived from a single qualifying study (Guignard, Bittner *et al*. 1981). However, the magnitude of this effect is lower than that observed for the perceptual-motor task categories.

The influence of WBV on performance was also considered with respect to other moderating variables. Performance in tasks emphasising accuracy of response were found to be degraded more than those emphasising the speed of response. Though these effects are not surprising (given the potential of WBV to cause unintended movements to body parts used in sensory and response processes) it is valuable to highlight that a better understanding of this type of effect may be achieved in future research by employing tasks that allow a direct trade-off between speed and accuracy, rather than looking at these as separate effects.

With regard to the characteristics of the vibration itself, it was found that performance decrements were larger for higher levels of frequency and intensity when compared against lower levels of these factors. The former is perhaps a little surprising given that performance is generally expected to degrade more at lower frequencies.

Though the International Standardisation Organization (ISO) no longer offers the guidance with respect to performance variation (ISO, 1997), previous versions (e.g. ISO, 1985) did provide Fatigue Decreased Proficiency (FDP) thresholds, which indicated that the effect of vibration frequency was moderated by the axis of vibration. Hence, our results may be a function of the present median split, and the moderating effects of axis of vibration. Of course, other factors are capable of moderating the frequency effect on performance, such as the axes of movements necessary for the performance response and the resonance frequencies of body parts used in the performance. However, resolution of these issues awaits further empirical examination. Although the present results are somewhat sparse, they do represent a synthesis of the available data and it is important to recognise that the exact mechanisms underpinning WBV frequency effects are not yet fully understood (Griffin, 2004). The moderating effects of WBV intensity imply that the greater the magnitude of stress, the greater the level of disruption. This is a general pattern of findings that is consistent with almost all theoretical constructs in stress (Matthews, Davies, *et al*. 2000; Hancock & Desmond, 2001). These findings also accord with those of Mansfield, Holmlund *et al*. (2000), who confirmed that higher magnitudes imposed a greater threat than low magnitudes, though the data were with reference to comfort, rather than performance.

Performance capacity was found to decrease as a function of duration alone, although the magnitude of this decrement function was moderated by the intensity of the WBV stimulus. Thus, the longer the performer is exposed to the higher intensities of WBV, then the greater the performance decrement. The current ISO standard (ISO, 1997) offers no guidelines concerning performance since it is thought to depend on several ergonomic factors.

4.1 Current state of knowledge

Overall, two main themes emerge from the present results. First, is evident that the current consensus is one of induced performance decrement; performance being degraded in all categories examined with the most deleterious associated with perceptual tasks. Second, this conclusion is based upon relatively few valid empirical investigations. This represents an example of what Laughery (1993) has referred to as the '*everybody knows*' problem, and one that Poulton (1976) has also articulated, especially in relation to stress effects. That is, since the general expectation is that vibration acts to degrade performance, few institutions or researchers are motivated to support or conduct extensive research to confirm an outcome that they think they already know. This is a potentially dangerous situation since, quite often, it is those very forms of consensus assumption that can prove fundamentally wrong. In the present context, our assumption of general degradation represents a conservative assertion but it is one that may cause the expenditure of unnecessary resources on mitigation technologies since decrement may not be ubiquitous under all combinatorial conditions.

In a recent review of WBV effects, Griffin (2004) concluded that though some general mechanisms are understood, there are no detailed models of WBV effects because we do not fully understand the relationships among the many moderating factors. Griffin asserted that attempts to describe the extent of performance interference under WBV are not as valuable as attempts to uncover the reasons for the observed interference. That is, to paraphrase Griffin, work undertaken to examine *whether* or *how* performance is degraded is less important than work that seeks to explain *why* it is degraded. Griffin did however emphasise that an effort that seeks to specify the relative importance of the various factors would be highly valuable (Griffin, 2004). Though the present effort does not answer this call for a mechanisms-led approach, the results of the present work go some of the way to detailing the relative influence different moderators has on the WBV – performance relationship. In detailing the current state of the research, the present paper also highlights the gaps in the experimental literature. We hope that our efforts can stimulate further work on two fronts. Firstly, it is hoped that the quantitative (statistical) and qualitative (highlighting of areas with a lack of research) findings from the present meta-analysis can provide an impetus and direction for further empirical investigations. Secondly, we suggest the adoption of a different perspective from which to tackle WBV-performance research.

4.2 The adaptive human performer

Understanding the specific WBV mechanisms stress on human may be better understood if future empirical studies derive from a 'top-down' theory-driven perspective. Humans are active agents in their world, and are capable of adapting to environments when motivated to do so (Hockey, 1997; Teichner, 1968). A logical step therefore is to recognise that the understanding of stressor-performance relationships needs to place the human at the centre of the assessment methods. Two candidate models are considered below with a view to them being used to inform the choice of factors to manipulate and the variables to measure, in addition to being used as a framework on which to examine the outcome results. Hence, though the present paper does not suggest possible reasons for *why* performance is disrupted under WBV, we propose candidate theoretical frameworks that may be used to uncover these mechanisms. We illustrate this by offering speculative explanations for the dose-response effect observed in the present data, in addition to WBV effects on cognitive performance, an area that has suffered from limited empirical examination (Sherwood, 1987),

4.2.1 The maximal adaptability model (Hancock & Warm, 1989).

An example of how a theory or model can guide the selection of factors to manipulate is the adoption of Hancock and Warm's (1989) maximal adaptability model. Hancock and Warm described three facets of stress, which they referred to as the 'trinity of stress.' Stressors such as vibration would constitute one of the 'input' factors along with the characteristics of the tasks (e.g. display organisation). It is to this facet of stress that the current meta-analytic results are most relevant. In regard to the effect of stress on behavioural and physiological adaptation, a central feature of the Hancock and Warm (1989) model is that under most environmental conditions individuals adapt effectively to

the input disturbance and maintain performance capacity. A second feature is that adaptation occurs at multiple levels, particularly the physiological, behavioural (performance) and subjective/affective levels. These levels are represented as a nested structure (see Figure 2), such that as the stress on the individual increases, by increased intensity, duration, or both, adaptation progressively fails. The first failure occurs in subjective state, followed by performance, with physiological failure as the ultimate failure in adaptation. The threshold between stable adaptation and instability (adaptation failure) has been observed in high stress environments, including physiological failure manifested as unconsciousness (Harris, Hancock, *et al*. 2005).

[Insert Figure 2 about here].

This model is unique in that the input stress from the environment extends beyond the physical or social environment to include the task itself. Indeed, Hancock and Warm (1989) argued that the task is often the most proximal form of stress in many real-world circumstances. Thus, the task characteristics are important determinants of adaptive state. This is represented in Figure 3 by two base axes reflecting the spatial and temporal characteristics of the task. Information structure (the spatial dimension) refers to how the task elements are organised, including workload and task complexity. The temporal dimension is represented as information rate. Together these dimensions can be used to form a vector (see Figure 3) which serves to identify an individual's adaptive state (i.e. point on the surface). Environmental inputs such as vibration could be an additional term in the vector (suggesting an n-dimensional model) or it could be used as part of the input determining the position of a task along the two existing dimensions shown in the figure. While the current results provide quantitative estimates of vibration effects under different conditions, more research is needed to determine how such information can be integrated within the model shown in Figure 3, and the relations among different sources of stress (i.e. multiple physical stressors and task-based stress; cf, Broadbent, 1971).

[Insert Figure 3 about here].

If the data from Table 2 are considered in light of the maximal adaptability model, it is apparent that, as one might expect, the combination of stress intensity and duration exert multiplicative rather than additive effects. Performance degradation therefore reflects something beyond a simple additive effect. In spite of the arguments that may be made about the present median-split technique to derive high and low levels of intensity and duration, the outcome implies a non-linear change in performance deterioration (primarily expressed here as the duration effect). This apparent synergistic effect of intensity and duration moderators is therefore an area worthy of further empirical investigation.

The maximal adaptability model can be used as a guide for WBV research. For instance, the representation of environmental and task characteristics as a vector (see Figure 3) indicates that one might systematically vary vibration characteristics (e.g. frequency, intensity, duration of exposure) and examine these at multiple levels of task characteristics (e.g. simple vs. complex information structure, high vs. low information

rate). The multiplicative effects indicated in the Hancock and Warm (1989) model can then be evaluated empirically. Of particular interest is the interaction between intensity (the abscissa in Figure 2) and duration of exposure. Examination of these interactive effects would fill a gap in stress research, which has traditionally neglected multiplicative effects (see Broadbent, 1971, for an early discussion of this issue). In addition, according to Hancock and Warm (1989) individuals exert effort to adapt to stress by narrowing of attention (see also Easterbrook, 1959), but such efforts become maladaptive when task relevant cues are excluded from attention and the duration of exposure (and effort to cope with the stress) exceed the capacity of the organism to maintain a stable level of adaptation. Thus, future investigation should examine the effect of vibration on mechanisms of selective and sustained attention, with specific emphasis on cognitive resource allocation. Finally, the maximal adaptability model recognises that there are multiple levels of stress response (cf, Matthews, 2001). Thus, as stress increases, failures to adapt occur progressively from subjective comfort to physiological failure. Future research should empirically test this progression to the extent possible by examining the changes in cognitive state and performance as a function of time on task and exposure to WBV. Establishing the progression of multiple stress responses would also allow more precise identification of the transition points between levels of adaptation (e.g. between the comfort and psychological zones in Figure 2) and the thresholds for adaptation failures.

4.2.2 The compensatory control model (Hockey, 1997)

Griffin (2004) proposed that one reason for the limited understanding of the mechanisms underpinning WBV effects may be the use of methods that do not capture the effects to a satisfactory degree (e.g. the use of simple reaction time measures of cognitive performance). This reasoning may be extended to actual *concepts*, with the definitions of the concepts not being sufficiently broad. Hence, certain WBV effects on performance may remain hidden from the experimenter, through the concept not being assessed to its full potential. A key example of this is the view of *performance* adopted in most empirical WBV research.

Hockey's Compensatory Control Model (CCM; Hockey, 1997) points out that simple methods may not be sufficient to capture stressor effects, as the performer may choose to 'protect' the level of observable performance through the application of increased effort or a change in strategy (also see Kahneman, 1973; Teichner, 1968). Hence, performance can be maintained under high levels of both environmental stress and task demands. A framework is therefore required that can identify stressor effects on a different level. This is achieved through the recognition that though performance may be protected, it is at a cost to the performer on other levels. Uncovering these 'latent' effects may therefore identify when a performer may be in a high-risk 'strain' state, where performance may be in risk of breakdown even though observable levels may appear unthreatened.

A key tenet underpinning Hockey's model is the importance of recognising the biological and motivational context in which performance takes place. Performance has to compete with other motivational goals (e.g. seeking rest), and as such, goals may change over the duration of performance. When performance under high demand is sustained,

performance goals are maintained in focal attention through the mobilisation of effort. As effort is a key moderating factor, it can be seen why observed performance levels under stress may degrade, improve, or stay the same – an observation also recognised by Griffin (2004). In order to identify strained performance, CCM proposes that a broad definition of performance should be used – one that considers not only performance effectiveness but also performance *efficiency* (i.e. a perspective that recognises the costs of maintaining performance).

To assess performance efficiency, Hockey proposed four types of latent performance decrement that may occur. First, increased effort levels are biased towards the protection of high-priority task goals and therefore may be reflected in the relative neglect of lower-priority tasks. Though secondary tasks are less critical to overall performance levels, the decrements introduce risk into the system. This decrement is a robust phenomenon, and was used initially to assess processing capacities in dual-task methodologies (e.g. Hancock & Meshkati, 1988; Moray, 1979). Second, performers under stress may adopt less resource-intensive, but more reactionary strategies (e.g. Sperandio, 1978). Third, the protection of performance is at a cost to the performer, with increases in subjective levels of effort expenditure, fatigue, and anxiety being reported, in addition to levels of psychophysiological sympathetic dominance. Finally, fatigue after-effects may be seen following prolonged effortful engagement. After-effects are considered to be the most valid test of fatigue levels (see Broadbent, 1979; Holding, 1983). In these situations, the tired operator is more likely to adopt low-effort (and therefore more risky) performance strategies, through being *unable* or *unwilling* to invest further effort.

Several other methodological considerations are emphasised by Hockey (1997), such as the importance of training participants when investigating stressor effects on performance. This yields increased ecological validity if the results are to be generalised to trained workers, and minimises the threat of learning effects masking any effect the stressor may have, as also recognised by Griffin (2004). As trained participants are more likely to seek to protect performance under high stress (as workers in operational work environments would), this further emphasises the importance of assessing latent decrements in order to evaluate threat to system performance.

In order to capture these hidden effects of stress on performance, WBV researchers are advised to adopt the convergent method approach advocated by CCM. As performers can protect the most salient aspects of performance but at a cost to less important tasks, the use of a multi-level performance tasks allow this trade-off to be captured. Self-report methods allow the assessment of regulatory activities. For instance, the measurement of subjective effort, anxiety and fatigue can highlight changes in performance strategy and the costs of compensatory behaviour (psychophysiological measures can also be used to triangulate the data yet further). After-effects of WBV exposure can be captured through the use of probe tasks following the main performance task.

Though CCM has not been used as a framework on which to base the empirical investigation of WBV effects, it has been used to investigate other stressors. For instance, Hockey, Wastell, *et al*. (1998) manipulated the levels of sleep deprivation and interface

control as stressors. He found that overt (primary task) performance levels were maintained under all levels of stress, with the only performance effects found for a secondary task under the most demanding of conditions. The results also showed that the protection of performance was achieved at a cost to increased levels of effort and fatigue. Conway (2005) found strong support for CCM over a series of five experiments in which the levels of task demand and environmental stressors (continuous broadband noise) were manipulated. Again, performers were able to protect primary task performance levels, though at a cost to secondary task performance levels. Increased effort mobilisation was found under high demand, resulting in high levels of strain-related variables (fatigue and anxiety).

Though the FDP guidelines were omitted from the latest ISO standards (ISO, 1997), viewing the dose-response effects found in the present paper with respect to CCM may actually offer speculative support for the role of *fatigue* in WBV effects on performance. It is important to be clear on what 'fatigue' is considered to be however to avoid the confusion that is common with this often-used but poorly-defined concept. Fatigue in CCM is considered to be an adaptive response to the interaction with task and environmental demands, and is defined as the resistance to the investment of further effort. Hence, the objective levels of task demand or WBV do not cause fatigue; rather, it is the performer's attempts to adapt to the imposed demands. Griffin (2004) previously questioned the proposed FDP effect on the grounds of it being more likely to be a factor of motivation levels. However, we and others (e.g. Hockey, 1997) stress that the two concepts, motivation and fatigue, are intimately connected, with reductions in task-directed effort being a possible mechanism for WBV dose-response effects. A tentative link can be made here between reductions in comfort and (short term) health, as proposed by ISO (1997) and possible reductions in performance. As the individual becomes more uncomfortable or becomes aware of health issues, then goals (and attention) can shift slightly away from performance in an attempt to restore prior levels of comfort and/or health. For instance, the operator may think about, or may take, a short break away from the task.

It is therefore emphasised that for WBV mechanisms on performance to be understood to a greater degree, it is importance to capture the broad range of effects on the operator's functional state, as Hockey's CCM proposes. The model provides a framework on which to conduct empirical assessments of WBV-performance relationships, it provides hypotheses to be tested, and can be used to interpret the results gained. In doing so, Hockey's CCM answers Sherwood's (1987) appeal for such a framework.

5 Conclusions and Recommendations

The study of WBV effects has been a traditional concern of the physical ergonomist who is tasked with the protection of individuals in the workplace and beyond. WBV is an evident source of physical stress but the way in which cognitive and perceptual-motor performance is influenced has been a surprisingly underserved enterprise. Here, we have presented a quantitative assessment of the present state of knowledge, but have also advocated the integration of WBV effects into the wider realm of stress theories. This strategy has the advantage of cross-referencing insights derived from such general

formulations to guide future research on all vibration effects. Interestingly, WBV does have direct effects on tasks of almost purely cognitive content. This implies therefore that as well as the manifest disturbances to the physical surface of support, vibration exerts a more indirect effect on human cognition, perception, and motor response.

While it is moral imperative of the ergonomist to protect individuals from physical harm, the evolution of the workplace towards that of an information marketplace means that issues such as health and comfort can no longer dominate the vibration landscape. It is fundamentally immaterial if an individual exposed to WBV is both healthy and comfortable but still making such egregious performance errors that they endanger themselves, their co-workers and the greater society served by the complex technological systems that they are in control.

Acknowledgement

The research reported in this document was performed in connection with contract DAAD19-01-C-0065 with the U.S. Army Research Laboratory. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory, or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. The authors would like to thank Mr. John Lockett and Ms. Sue Archer for providing administration and technical direction, in addition to comments on the present work.

This work was facilitated by the Department of Defense Multi disciplinary University Research Initiative (MURI) program, P.A. Hancock, *Principal Investigator*, administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

We would also like to acknowledge Ms. Jennifer Ross and Mrs. Bonnie Saxton for their assistance in the collection and coding of articles for the meta-analysis. We would like to thank Dr. Bob Kennedy and Dr. John Guignard for their expertise and commentary on our project. Finally, we wish to thank the unknown Reviewers of this work whose insightful comments were significantly helpful in revising this paper.

References (* indicates papers included in the formal meta-analysis)

ARNBERG, P. W., BENNERHULT, O. & EBERHARDT, J. L., 1990, Sleep disturbances caused by vibrations from heavy road traffic. *Journal of the Acoustical Society of America,* 88**,** pp. 1486-1493.

BADDELEY, A. D., 1986, *Working Memory*. (Oxford, Oxford University Press).

BROADBENT, D. E., 1971, *Decision and Stress.* (London, Academic Press).

*CLARKE, N. P., TAUB, H., SCHERER, H. F. & TEMPLE, W. E., 1965, *Preliminary study of dial reading performance during sustained acceleration and vibration*. (Dayton, Ohio, Air Force Research Lab).

*COHEN, H. H., WASSERMAN, D. E. & HORNUNG, R. W., 1977, Human performance and transmissibility under sinusoidal and mixed vertical vibration. *Ergonomics,* 20**,** pp. 207-216.

COHEN, J., 1988, *Statistical Power Analyses for the Behavioral Sciences, (*Hillsdale, NJ, Lawrence Erlbaum Associates).

COHEN, S., 1980, After effects of stress on human performance and social behavior: A review of research and theory. *Psychological Bulletin,* 88**,** pp. 82-108.

CONWAY, G. E., 2005, *Effects of workload, effort and fatigue on complex performance: application and tests of compensatory control theory*. Unpublished Ph.D. thesis. (Leeds, UK, University of Leeds).

*DEAN, R. D., FARRELL, R. J. & HITT, J. D., 1969, Effect of vibration on the operation of decimal input devices. *Human Factors,* 11**,** pp. 257-272.

DRISKELL, J. E. & MULLEN, B., 2005, The efficacy of naps as a fatigue countermeasure: A meta-analytic integration. *Human Factors,* 47**,** pp. 360-377.

EASTERBROOK, J. A., 1959, The effect of emotion on cue utilisation and the organisation of behaviour. *Psychological Review,* 66**,** pp. 183-201.

FRASER, W., SVENSSON, E., GRANDT, M., HOCKEY, G. R. J., BALKIN, T., BEAUMONT, M., MAMIMORI, G., KAUTZ, M., BELENKY, G. L., WESENSTEN, N. & SCHLEGEL, R. E., 2004, Risk Factors. In RTO Human Factors and Medicine Panel (HFM) Task Group (HFM-056/TG-008). *Operator Functional State Assessment* (Technical Report RTO-TR-HFM-104). *(*Neuilly-sur-Seine Cedex, France, Research and Technology Organisation, North Atlantic Treaty Organisation).

*GAUTHIER, G. M., ROLL, J. P., MARTIN, B. & HARLAY, F., 1981, Effects of whole-body vibrations on sensory motor system performance in man. *Aviation, Space & Environmental Medicine,* 52**,** pp. 473-479.

GRIFFIN, M. J., 1992, Vibration. In *Handbook of Human Performance*, A. P. SMITH, & D. M. JONES, (Eds) (London, Academic Press).

GRIFFIN, M. J., 2004, *Handbook of Human Vibration,* (London, Elsevier).

*GUIGNARD, J. C., BITTNER, A. C. & CARTER, R. C., 1981, *Methodological investigation of vibration effects of performance of three tasks*, (New Orleans, Louisiana, Naval Biodynamics Laboratory).

HANCOCK, P. A. & DESMOND, P. A., 2001, *Stress, Workload and Fatigue,* (Mahwah, NJ, Lawrence Erlbaum Associates).

HANCOCK, P. A. & MESHKATI, N. (Eds.) 1988, *Human Mental Workload,* (Amsterdam, North-Holland).

HANCOCK, P. A. & WARM, J. S., 1989, A dynamic model of stress and sustained attention. *Human Factors,* 31**,** pp. 519-537.

HARRIS, W. C., HANCOCK, P. A. & HARRIS, S. C., 2005, Information processing changes following extended stress. *Military Psychology,* 17**,** pp. 115-128.

HEDGES, L. V. & OLKIN, I., 1985, *Statistical Methods for Meta-Analysis,* (New York, Academic Press).

HEDGES, L. V., SHYMANSKY, J. A. & WOODWORTH, G., 1989, *A Practical Guide to Modern Methods of Meta-Analysis,* (Washington, DC, National Science Teachers Association).

HOCKEY, G. R. J. (Ed.), 1983, *Stress and Human Performance,* (Chichester, John Wiley & Sons).

HOCKEY, G. R. J., 1997, Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology,* 45**,** pp. 73-93.

HOCKEY, G. R. J. & HAMILTON, P., 1983, The cognitive patterning of stress states. In *Stress and Human Performance,* G. R. J. HOCKEY, (Ed.) (Chichester, John Wiley & Sons).

HOCKEY, G. R. J., WASTELL, D. & SAUER, J., 1998, Effects of sleep deprivation and user interface on complex performance: A multilevel analysis of compensatory control. *Human Factors,* 40**,** pp. 233-253.

HUNT, M. M., 1997, *How Science Takes Stock: The Story of Meta-Analysis,* (New York, Russell Sage Foundation).

HUNTER, J. E. & SCHMIDT, F. L., 2004, *Methods of Meta-analysis: Correcting for Error and Bias in Research Findings,* (Thousand Oaks, CA, Sage).

INTERNATIONAL ORGANIZATION for STANDARDISATION, 1985, *Evaluation of human exposure to whole-body vibration: Part 1 – general requirements*. *ISO 2631/1* (Geneva, International Organization for Standardization)

INTERNATIONAL ORGANIZATION for STANDARDISATION, 1997, *Mechanical vibration and shock: Evaluation of human exposure to whole-body vibration: part 1 – general requirements*. ISO 2631-1 (Geneva, International Organization for Standardization)

KAHNEMAN, D., 1973, *Attention and Effort,* (Englewood Cliffs, N.J., Prentice-Hall).

LAUGHERY, K. R., 1993, Everybody Knows - Or Do They? *Ergonomics in Design,* .3**,** 8.

LIPSEY, M. W., 2003, Those confounded moderators in meta-analysis. *Annals of the American Academy of Political and Social Science,* 587**,** pp. 69-81.

LIPSEY, M. W. & WILSON, D. B., 2001, *Practical Meta-Analysis,* (Thousand Oaks, CA, Sage Publications Inc).

*LOEB, M., 1954, *A preliminary investigation of the effects of whole-body vibration and noise*. (Fort Knox, KY., Army Medical Research Laboratory).

MANSFIELD, N. J., 2004, *Human Response to Vibration,*(Boca Raton, FL., CRC Press).

MANSFIELD, N. J., HOLMLUND, P. & LUNDSTRÖM, R., 2000, Comparison of subjective responses to vibration and shock with standard analysis methods and absorbed power. *Journal of Sound and Vibration,* 230**,** pp. 477-491.

MATTHEWS, G., 2001, Levels of transaction: A cognitive science framework for operator stress. In *Stress, Workload, and Fatigue,* P. A. HANCOCK, & P. A. DESMOND (Eds.) (Mahwah, NJ, Lawrence Erlbaum Associates).

MATTHEWS, G., DAVIES, D. R., WESTERMAN, S. J. & STAMMERS, R. B., 2000, *Human Performance: Cognition, Stress and Individual Differences* (Hove, Sussex, Psychology Press).

*MCLEOD, R. W. & GRIFFIN, M. J., 1993, Effects of duration and vibration on performance of continuous manual control task. *Ergonomics,* 36**,** pp. 645-659.

MEISTER, A., BRAUER, D., KUREROV, N. N., METZ, A. M., MUCKE, R., ROTHE, R., SEIDEL, H., STAROZUK, I. A. & SUVOROV, G. A., 1984, Evaluation of responses to broad-band whole-body vibration. *Ergonomics,* 27**,** pp. 959-980.

MORAY, N. (Ed.), 1979, *Mental Workload: Its Theory and Measurement,* (New York, Plenum Press).

MORRIS, S. B. & DESHON, R. P., 2002, Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods,* 7**,** pp. 105-125.

O' DONNELL, R. D. & EGGEMEIER, F. T., 1986, Workload assessment methodology. In *Handbook of Perception and Human Performance,* K. R. BOFF, L. KAUFMAN, & J. P. THOMAS, (Eds.) (New York, John Wiley & Sons).

PEACOCK, B., CHASE, G., FOX, R. & PARKER, R., 2005, Unwanted Energy - Vibration. *Ergonomics in Design,* 13**,** 4, pp. 27-28, 31.

POULTON, E. C., 1976, Arousing environmental stresses can improve performance, whatever people say. *Aviation, Space, and Environmental Medicine,* 47**,** pp. 1193-1204.

RAMAZZINI, B., 1713, *De Morbis Artificum (Diseases of workers),* (New York, Hafner Publishing Co).

*RAO, B. K. N. & ASHLEY, C., 1974, Effect of whole-body low frequency random vertical vibration on vigilance task. *Journal of Sound and Vibration,* 33**,** pp. 119-125.

SHERWOOD, N., 1987, Vibration and cognition: Joining the boundaries. *United Kingdom informal group meeting on human response to vibration.* (Royal Military College of Science, Shrivenham).

*SHOENBERGER, R. W., 1967, Effects of vibration on complex psychomotor performance. *Aerospace Medicine,* 38**,** pp. 1264-1269.

*SHOENBERGER, R. W. & WILBURN, D. L., 1973, *Tracking performance during whole-body vibration with side-mounted and center-mounted control sticks*. (Dayton, Ohio, Wright Patterson AFB).

SOMMER, B., 1987 The file drawer effect and publication rates in menstrual cycle research. *Psychology of Women Quarterly,* 11**,** pp. 233-242.

SPERANDIO, J. C., 1971, Variation of operators' strategies and regulating effects on workload. *Ergonomics,* 14**,** pp. 571-577.

SUZUKI, H., 1998, Effects of the range and frequency of vibrations on the momentary riding comfort evaluation of a railway vehicle. *Japanese Psychological Research,* 40**,** pp. 156-165.

TEICHNER, W. H., 1968, Interaction of behavioral and physiological stress reactions. *Psychological Review,* 75**,** pp. 271-291.

WARM, J. S. (Ed.), 1984, *Sustained Attention and Human Performance,* (Chichester, John Wiley & Sons).

*WEISZ, A. Z., GODDARD, C. & ALLEN, R. W., 1965, *Human performance under random and sinusoidal vibration*. (Dayton, Ohio, Aerospace Medical Research Laboratory, Wright-Patterson AFB).

WICKENS, C. D. & HOLLANDS, J. G., 2000, *Engineering Psychology and Human Performance,* (Upper Saddle River, New Jersey, Prentice-Hall).

Biographies



Dr Gareth Conway is a Research Associate in the MIT^2 lab, within the Institute of Simulation and Training and Department of Psychology, at University of Central Florida. Dr Conway lead the lab's efforts on updating the knowledge concerning stressor influences within the US Army's IMPRINT performance modelling tool through the use of meta-analytic methods. He received his Ph.D. in 2005 from the Institute of Psychological Sciences, University of Leeds, where he investigated compensatory performance and operator functional state assessment in complex tasks, under the supervision of Professor Bob Hockey. The thesis was awarded the 2006 George E. Briggs Dissertation Award from Division 21 (Applied Experimental and Engineering Psychology) of the American Psychological Association.



Dr James Szalma is an assistant professor in the psychology department at the University of Central Florida. He received his Ph.D. in 1999 from the University of Cincinnati. Dr. Szalma's current research efforts are aimed at testing theoretical models of workload, stress and performance on signal detection tasks, particularly those involving sustained attention and monitoring. As part of this work, he is investigating individual differences in personality and their relation to performance, workload, stress, and coping strategies. Dr. Szalma is also conducting studies to empirically test a novel modification of Signal Detection Theory, Fuzzy Signal Detection Theory, as well as measures that serve as alternatives to traditional SDT.



Dr Peter Hancock is Provost's Distinguished Research Professor in the Department of Psychology Institute for Simulation and Training at the University of Central Florida. He is the author of over three hundred refereed scientific articles and publications. He has been continuously funded by extramural sources for every year of his professional career, including support from NASA, NIH, NIA, FAA, FHWA, and the US Navy, and Air Force. He is the Principal Investigator on the Multi-University Research Initiative (MURI) on stress and soldier performance. In 2000, he was awarded the Sir Frederic Bartlett Medal of the Ergonomics Society of Great Britain for lifetime achievement. He was recently awarded a Jastrzebowski Medal of the Polish Ergonomics Society for the contributions to world ergonomics.

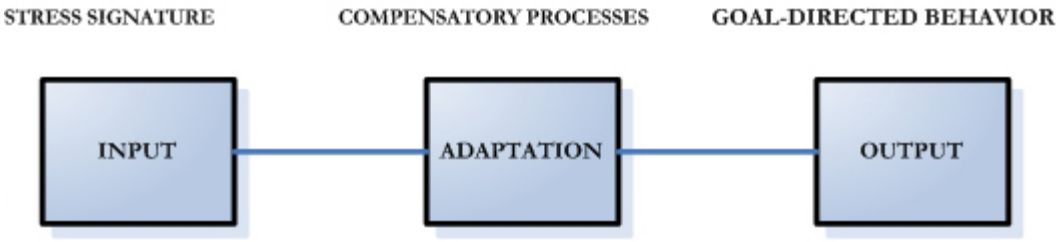


Figure 1. The 'trinity of stress' (Hancock & Warm, 1989). A tripartite descriptive framework for describing the environmental origin of stress (input), its representation as a direct pattern of adaptive, regulatory responses (adaptation) and its manifestation in disturbance to on-going performance capacity (output).

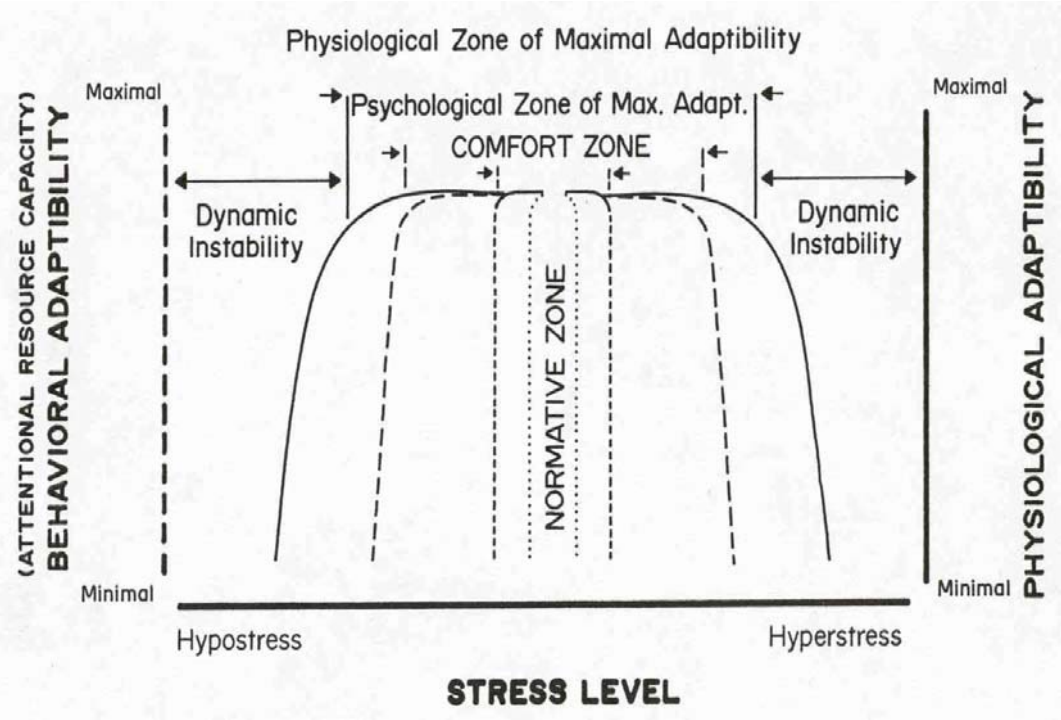


Figure 2. The extended-U relationship between stress level and response capacity, from the Maximal Adaptability Model (Hancock & Warm, 1989).

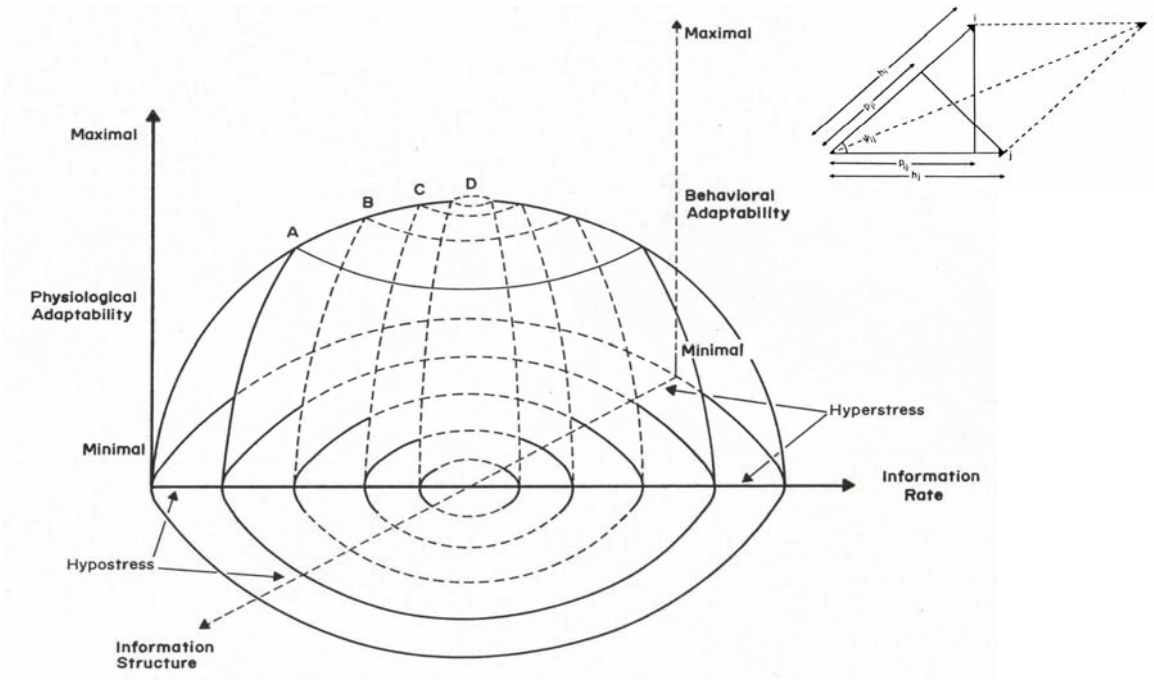


Figure 3. The description given in Figure 2 is expanded into a three-dimensional representation by parsing the base 'hypostress-hyperstress' axis into its two component elements. These divisions are compose of information rate (the temporal axis) and information structure (the spatial axis).

Table 1. Formal Whole-Body Vibration Meta-Analysis Results

Analysis	k	*g*	s_d^2	s_e^2	95%CI (var(e))
Global	13	-0.95	1.14	0.75	0.47
Task					
Perceptual	4	-1.79	2.02	2.96	1.69
Cognitive	1	-0.52			
Fine Motor Continuous	8	-0.89	1.47	0.74	0.60
Fine Motor Discrete	3	-0.84	0.66	0.5	0.80
Dependent variable type					
Accuracy	9	-1.11	1.76	1.07	0.68
RT	5	-0.38	0.15	0.34	0.51
Duration					
Low	10	-0.87	0.94	0.48	0.43
High	3	-2.6	0.53	9.91	3.56
Intensity					
low	7	-0.67	1.41	0.4	0.47
high	9	-0.82	0.63	0.75	0.57
Frequency					
low	10	-0.65	0.73	0.5	0.44
high	5	-1.98	0.97	1.66	1.13
Duration by Intensity					
low duration, low intensity	5	-0.4	1.35	0.25	0.44
low duration, high intensity	8	-0.8	0.43	0.53	0.50
high duration, low intensity	2	-2.34	1.42	3.37	2.54
high duration, high intensity	1	-3.84			

Table 2. Duration by intensity interaction effect sizes

Duration			
Short		Long	
Low intensity	High intensity	Low intensity	High intensity
-0.4	-0.8	-2.34	-3.84

THE EFFECTS OF WHOLE-BODY VIBRATION ON HUMAN PERFORMANCE: A META-ANALYTIC EXAMINATION

G.E. Conway, J.L. Szalma, B.M. Saxton., J.M. Ross, and P.A. Hancock

MIT² Laboratory, 3100 Technology Parkway, Suite 337,
University of Central Florida, Orlando, FL, 32826, USA.

Whole-body vibration exerts a substantive influence in many work environments. The primary objective of the present paper was to ascertain the effect of whole-body vibration and identify those moderating variables that influence the degree to which human performance is affected. A comprehensive meta-analysis was conducted which quantified the existing research evidence. Following a screening process of the collected literature, a total of 224 papers and reports were identified for analysis. From these papers, 115 effect sizes were derived from 13 experiments which survived the screening procedure. Results indicate that vibration acts to degrade the majority of goal-related activities, especially those that rely on visual perception and fine motor control. Gaps in the extant research literature are identified and suggestions offered with regard to a more theoretically-driven approach to testing stressor effects on human performance.

INTRODUCTION

The problem of vibration in the work environment is an issue that has long been of concern (e.g., Ramazzini, 1713). A pervasive form of the stressor is whole body vibration (WBV), which manifests when vibrating surfaces support the human performer. Hence, the typical mechanism of vibration disruption to performance is direct, through the mechanical perturbation when the vibration is transmitted to and dissipated within the body. Vibration therefore influences several aspects of performance including sensory, and response processes, in addition to more latent effects on information processing and adaptive processes (the individuals' attempts to cope with their ambient environment). Though these effects are recognized, the exact mechanisms underpinning the specific effects have proved difficult to identify (Griffin, 2004). WBV exists in any number of work realms but particularly in those that require the concomitant use of both *transportation* and *information* systems, such as using high technology interfaces in vehicles. As such, the implications of vibration effects on performance should be of great interest to human factors professionals.

Purpose of the Current Work

Here we provide a comprehensive, quantitative analysis of all available, existing scientific information that we identified concerning the influence of whole-body vibration on human performance. We have investigated these effects using meta-analytic procedures (see Hedges & Olkin, 1985; Hunt, 1997; Hunter & Schmidt, 2004). The screening process resulted in the qualification of thirteen studies for inclusion. These emerged from a total of eleven different sources (i.e., articles, dissertations, technical reports, etc.), and resulted in one-hundred and fifteen effect sizes.

METHOD

Literature Accumulation

To collect the input values for the WBV meta-analysis, an exhaustive literature search was performed using the following search term combinations as a primary screen: 'whole body', 'vibration', 'performance', 'cognition', 'motor', and 'vigilance'. Additional articles were collected by surveying the reference lists from those already available and by retrospectively examining article citations through Science Citation Index® (SCI).

Criteria for Inclusion

Each study had to report an empirical examination of vibration stress in which the experimental manipulation employed an application of WBV, had to address the issue of direct WBV effects on performance, was required to include a

control group for comparison purposes, had to report at least one measure of performance, and had to include sufficient information regarding performance results to determine effect size estimates. Thirteen primary studies were subsequently accepted for use in the meta-analysis.

The Calculation of Effect Sizes

Effect sizes for this study were the standardized mean difference between the experimental and the control conditions, often referred to as Hedge's *g* (Hedges & Olkin, 1985). When means and/or standard deviations were unavailable the effect sizes were computed using inferential statistics (e.g., *t*-ratios) or sums of squares/mean squares (e.g., from ANOVA tables) using the equations presented by Lipsey and Wilson (2001).

Estimation of Variances

The variance associated with each effect size was calculated using equations provided by Morris and DeShon (2002). These were combined for estimates of sampling error variance (s_e^2) and the variance among the observed effect sizes (s_g^2) using procedures described by Hunter and Schmidt (2004). These estimates can be used to derive the variance due to differences in the population effect size.(s_δ^2). Thus,

$$s_\delta^2 = s_g^2 - s_e^2 \quad \text{---------- (1)}$$

A large s_δ^2 indicates that there is variability among the observed effect sizes that cannot be accounted for by sampling error, and that there are likely to be one or more variables moderating the magnitude of the effect in question (see Hunter & Schmidt, 2004, pg. 288). Note that if all of the variance in the effect sizes were accounted for by sampling error, then $s_\delta^2 = 0$. Another possibility is that the population effect size itself varies randomly (i.e., a random effects model, see Hedges & Vevea, 1998).

META-ANALYTIC RESULTS

In respect of the present results, the first observation concerns the rejection rate of studies that failed to meet the selection criteria. The collective survey of 224 studies and reports generated 13 useable studies (from only 11 papers). The results of the outcome meta-analysis are presented in Table 1. The global analysis shows an effect size of -0.95, which represents a large overall performance decrement. Since it includes 13 total studies, this represents a supportable conclusion as a number of other meta-analyses have been reported with similar numbers of studies (e.g., Driskell & Mullen, 2005).

A major aim of the present work was to identify and quantitatively assess WBV-performance moderating factors. The first moderator variable, *task characteristics,* was assessed through categorising the dependent variables used. Four performance categories were identified - *perceptual processes, cognitive processes, continuous fine motor actions* and *discrete fine motor actions*. An example of perceptual tasks is vigilance/target detection, working memory and mathematical reasoning are both examples of cognitive tasks, a continuous fine motor task could be tracking in nature, and an example of a discrete fine motor task is that of flicking a switch on a console.

Table 1. Formal whole-body vibration meta-analytic results.

Analysis	k	*g*	S_g^2	s_e^2	95%CI (var(e))
Global	13	-0.95	1.14	0.75	0.47
Task					
Perceptual	4	-1.79	2.02	2.96	1.69
Cognitive	1	-0.52			
Continuous FM	8	-0.89	1.47	0.74	0.60
Discrete FM	3	-0.84	0.66	0.5	0.80

Note: FM = Fine motor

Table 1 shows a substantial, large effect of WBV on perceptual tasks, with all four contributing studies reporting effects reflecting considerable performance decrements under WBV. A medium effect size of WBV on cognitive performance was also found, though this result should be interpreted with caution as it was drawn from only one study. The effect of WBV on *continuous fine motor* performance was found to be similar to the effect on *discrete fine motor* performance, with both categories of fine motor performance demonstrating large degradation effects under WBV.

Table 2. Moderator/hierarchical whole-body vibration meta-analysis results.

Analysis	k	*g*	s_g^2	s_e^2	95%CI (var(e))
Dependent variable type					
Accuracy	9	-1.11	1.76	1.07	0.68
RT	5	-0.38	0.15	0.34	0.51
Duration					
Low	10	-0.87	0.94	0.48	0.43
High	3	-2.6	0.53	9.91	3.56
Intensity					
Low	7	-0.67	1.41	0.4	0.47
High	9	-0.82	0.63	0.75	0.57
Frequency					
Low	10	-0.65	0.73	0.5	0.44
High	5	-1.98	0.97	1.66	1.13
Duration by Intensity					
low duration, low intensity	5	-0.4	1.35	0.25	0.44
low duration, high intensity	8	-0.8	0.43	0.53	0.50
high duration, low intensity	2	-2.34	1.42	3.37	2.54
high duration, high intensity	1	-3.84			

WBV effects on performance were also parsed according to the aspect of performance measured. Table 2 shows a greater decrement of WBV on the *accuracy* of performance compared with the *speed* of performance. It should be noted that we considered only separate examinations of speed and accuracy, rather than trade-offs between the two.

In order to evaluate the differential effects of the WBV characteristics, the separate effects of *vibration frequency* and *vibration intensity* were assessed. In the absence of clear theoretical guidance, we adopted an empirically-based division, rendering the data bifid via the employment of a median split technique to create 'high' and 'low' categories for intensity and frequency. Vibration intensity under 0.2 RMSg was considered to be low and over this level, high. Low intensities were found to exert a smaller effect than high intensities, implying that the greater the magnitude of stress, the greater the disruption. Frequency values greater than 5 Hz were taken to represent high frequency WBV, and under 5 Hz, low frequency. High frequencies exerted a large negative effect on performance, while low frequencies exerted a moderate negative effect (k= 10, *g*= -0.65).

Table 3. Duration by intensity interaction effect sizes.

Duration			
Short		Long	
Low intensity	High intensity	Low intensity	High intensity
-0.4	-0.8	-2.34	-3.84

A key factor in distilling WBV effects is that of *exposure duration*, with the premise being that for any given combination of intensity and frequency, increases in exposure duration reduce operator performance accordingly. The present results were categorised in two groups, studies that exposed performers to less than 30 min WBV, and studies exceeding this duration. Shorter exposure times (<30 min) produced a large effect on performance levels while longer durations (>30 min) produced a much larger effect, though the latter must be interpreted with caution due to the large variation found in the individual effects contributing to the calculated effect size (s_e^2 = 9.91).

In order to further evaluate the ISO thresholds, a secondary analysis was carried out to determine the extent to which exposure duration moderates the intensity effects of WBV on performance (using the median splits for intensity as described above). When short exposure durations were assessed as a function of low and high intensity, small and moderate effect sizes were found respectively. When the long exposure durations were assessed, a substantial large effect was found for long duration/low intensity, and an even larger effect was found for the combination of long duration/high intensity. It is therefore apparent that performance decrements increase as duration and intensity increase.

DISCUSSION

WBV has a negative influence on performance, with effects moderated by the type of task being performed. Perceptual tasks suffered the largest decrements, followed by continuous and discrete fine motor tasks. Hence, sensory and motor response process suffered most under WBV, reinforcing the support for direct mechanical

interference being a general mechanism of performance disruption. Though a moderate negative effect was found for WBV influences on cognitive performance the effect size was derived from only one qualifying study. Hence, discussion of decrements in central (information processing) performance relative to those found for input (sensory) and output (fine motor response) is difficult at this time.

Performance accuracy-based tasks were found to be degraded more than those emphasising the speed of the response, and decrements were larger for higher levels of frequency and intensity when compared against lower levels. Though it is intuitive that the greater the level of the stressor (the magnitude here), the greater the level of performance disruption, the result for frequency is surprising given that performance is generally thought to suffer more at lower frequencies. However, the interpretation of this result is difficult as frequency effects are moderated by a further factor, the axis of vibration. Performance has been found to suffer most in the 1-2 Hz range when in the x- or y-axis, and in the 4-8 Hz range in the z-axis. Hence, the result may be a function of the median split being between the two ranges.

Greater exposure durations were found to result in larger performance decrements. This was found with respect to duration alone, and also when exposure duration was moderated by the intensity of the WBV stimulus. Thus, the longer the performer is exposed to the higher intensities of WBV, then the greater the performance decrement.

The Current State of Knowledge

Two main themes emerge from the present results. The current consensus is one of induced performance decrement, with the most deleterious associated with perceptual tasks. Though a large number of studies were collected, few were rigorous enough and/or supplied the information needed to contribute to the present analysis. Hence, that this conclusion is based upon relatively few valid empirical investigations may be an example of what Laughery (1993) termed the '*everybody knows*' problem - since the general expectation is that vibration degrades performance, few are motivated to support extensive research to confirm this.

Griffin (2004) recently highlighted a lack of supporting evidence for the moderating effect of exposure duration. Another area in which there is a paucity of studies is the effect of WBV on cognitive performance (see Sherwood, 1987). The acknowledgements of poor understanding in these two areas may provide the impetus to approach WBV effects from a slightly different perspective. A great deal of the empirical studies in this area have focused on the characteristics of the vibration stressor. However, this focus may be unsuitable for the understanding of the mechanisms underpinning WBV exposure duration effects, or the effects on information processing components of performance. Humans are active agents in their world and are capable of investing more effect or changing their strategy to protect important performance goals under stress, or letting levels slip when performance is deemed to be of low importance relative to other goals (e.g., conservation of attentional resources, or moving away from the source of the environmental stressor).

Understanding WBV effects on performance may therefore benefit from the adoption of a 'top-down' perspective, using established theories of human performance under stress to inform the choice of experimental factors and variables, and as a framework on which to examine the results. Hancock and Warm's (1989) maximal adaptability model is recommended for its consideration of factors to manipulate, whilst Hockey's (1997) Compensatory Control Model is recommended as a framework on which to assess WBV effects on performance. Both models focus on the one consistent factor in the stressor-performance relationship – the human; and also pay particular attention to the adaptive capacities of the performer.

ACKNOWLEDGEMENTS

This research was or supported by a contract through Micro Analysis and Design. Component Improvement for Soldier-Centered Tools #800505410/TO#54 from the Army Research Laboratory at Aberdeen Proving Ground, Maryland, P.A. Hancock and J.L. Szalma, *Principal Investigators*. The views expressed here are those of the authors and do not necessarily reflect the official policy or position of the named agencies. The authors would like to thank Mr. John Lockett and Ms. Sue Archer for providing administration and technical direction, in addition to comments on the present work.

This work was facilitated by the Department of Defense Multi disciplinary University Research Initiative (MURI) program, P.A. Hancock, *Principal Investigator*, administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, and Dr. Mike Drillings for providing administrative and technical direction for the Grant.

We would like to thank Dr. Bob Kennedy and Dr. John Guignard for their expertise and commentary on our project. Finally, we wish to thank the unknown Reviewers of this work.

REFERENCES

Driskell, J.E., & Mullen, B. (2005) The efficacy of naps as a fatigue countermeasure: A meta-analytic integration. *Human Factors, 47*, 360-377.

Griffin, M.J. (2004) *Handbook of human vibration,* London, Elsevier.

Hancock, P.A., & Warm, J.S. (1989) A dynamic model of stress and sustained attention. *Human Factors, 31*, 519-537.

Hedges, L.V., & Olkin, I. (1985) *Statistical methods for meta-analysis,* New York, Academic Press.

Hedges, L.V. & Vevea, J.L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3* (4), 486-504.

Hockey, G.R.J. (1997) Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*, 73-93.

Hunt, M.M. (1997) *How science takes stock: The story of meta-analysis,* New York, Russell Sage Foundation.

Hunter, J.E. & Schmidt, F.L. (2004) *Methods of meta-analysis: Correcting for error and bias in research findings,* Thousand Oaks, CA, Sage.

Laughery, K.R. (1993) Everybody Knows - Or Do They? *Ergonomics in Design., 3*, 8.

Lipsey, M.W. & Wilson, D.B. (2001) *Practical meta-analysis,* Thousand Oaks, CA, Sage Publications Inc.

Morris, S.B. & Deshon, R.P. (2002) Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105-125.

Ramazzini, B. (1713) *De Morbis Artificum (Diseases of workers).* (Translation by WC Wright, 1964). New York, Hafner Publishing Co

13

Peter A. Hancock and James L. Szalma

Stress and Neuroergonomics

Neuroergonomics has been defined as "the study of human brain function in relation to work and technology" (Parasuraman, 2003, p. 5). Given that stress is an aspect of many forms of work, it is natural that investigations of stress and its neurobehavioral aspects form a core topic in neuroergonomics. We begin this chapter with a brief overview of the major conceptual approaches to the study of stress, including a short historical account and a précis of the more recent stress theories (e.g., Hancock & Desmond, 2001). We discuss how an understanding of stress helps shape and direct new, emerging concepts in neuroergonomics. We then consider the issue of individual differences and the effects on stress in individuals, and ethical issues related to monitoring and mitigating stress in the workplace.

Concepts of Stress

In behavioral research, stress itself has traditionally been viewed as a source of disturbance arising from an individual's physical or social environment. However, since individuals do not react in exactly the same way to common conditions, it is now considered more appropriate to view stress in terms of each individual's response to his or her environment. This is the so-called transactional perspective (see Hockey, 1984, 1986; Lazarus & Folkman, 1984; Matthews, 2001; Wickens & Hollands, 2000). This has led to both continuing debate and some confusion, since some researchers continue to define stress in terms of the external stimuli involved (e.g., noise, temperature, vibration; see Elliott & Eisdorfer, 1982; Jones, 1984; Pilcher, Nadler, & Busch, 2002). Defining stress only in terms of the physical stimulus does not account for why the same stimulus induces different stress responses across individuals or within the same individual on different occasions (Hockey, 1986; Hockey & Hamilton, 1983; Matthews, 2001). Consideration of stimulus properties therefore provides an important but nevertheless incomplete understanding of stress effects. To more fully capture all the dimensions of stress, we have developed a "trinity of stress" model (Hancock & Warm, 1989), which includes (1) environmental stimulation as the input dimension; (2) the transactional perspective, emphasizing the individual response as the adaptation facet; and (3) most critically, an output performance level. We return to this description after a review of some theoretical background.

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

Arousal Theory

In contrast to the stimulus-driven view of stress, which might be considered a physics-based approach, the biologically based approaches define stress in terms of the physiological response patterns of the organism (Selye, 1976). In this vein, Cannon (1932) conceptualized stress in terms of autonomic nervous system activation triggered by a homeostatic challenge. According to Cannon, external conditions (e.g., temperature, noise, etc.) or internal deficiencies (e.g., low blood sugar, etc.) trigger deviations from homeostatic equilibrium. Such threats to equilibrium result in physiological responses aimed at countering that threat. These responses involve sympathetic activation of the adrenal medulla and the release of several hormones (see Asterita, 1985; Frankenhaeuser, 1986; for an early review see Dunbar, 1954). The response-based approach of Cannon was also championed by Selye (1976), who defined stress in terms of an orchestrated set of these bodily defense reactions against any form of noxious stimulation. This set of physiological reactions and processes he referred to as the *general adaptation syndrome*. Environmental objects or events that give rise to such responses were referred to as *stressors*. Within Selye's theory, physiological responses to stressors are general in character, since the set of responses is similar across different stressors and contexts.

Arousal theory is one of the most widely applied physiological response-based theories of stress and performance (Hebb, 1955). *Arousal level* is a hypothetical construct representing a nonspecific (general) indicator of the level of stimulation of the organism as a whole (Hockey, 1984). Arousal may be assessed using techniques such as electroencephalography (EEG) or indicators of autonomic nervous system activity such as the galvanic skin response (GSR) and heart rate. As a person becomes more aroused, the waveforms of the EEG increase in frequency and decrease in amplitude (see also chapter 2, this volume), and skin conductance and heart rate increase (see also chapter 14, this volume).

Within this framework, stress effects are observed under conditions that either overarouse (e.g., noise) or underarouse the individual (e.g., sleep deprivation; Hockey, 1984; McGrath, 1970). This approach assumes an inverted-U relationship between arousal and performance—the Yekes-Dodson law—such that the optimal level of performance is observed for midrange levels of arousal (Hebb, 1955). Stressors, such as noise or sleep loss, act by either increasing or decreasing the arousal level of the individual relative to the optimum level for a given task (Hockey & Hamilton, 1983). The optimum level is also postulated to be inversely related to the difficulty of the task (Hockey, 1984, 1986). A potential mechanism to account for this relation was first postulated by Easterbrook (1959), who indicated that emotional arousal restricts the utilization of the range of peripheral visual cues from the sensory environment, so that, under conditions of chronic or acute stress, peripheral stimuli are less likely to be processed than more centrally located cues. As attention narrows (i.e., as the number of cues attended to is reduced), performance capacity is preserved by the retention of focus on salient cues. Eventually, performance fails as stress increases and even salient cues become excluded. Hancock and Dirkin (1983) showed that the narrowing phenomenon was attentional rather than sensory in nature, since individuals narrowed to the source of greatest perceived salience wherever it appeared in the visual field. More recently, Hancock and Weaver (2005) showed that the narrowing phenomenon demonstrated in spatial perception is also evident in the temporal domain, as in the effects of stress on time perception (see also Szalma & Hancock, 2002).

There are several problems with the traditional arousal explanation of stress and performance. First, the different physiological indices of arousal often do not correlate well. In the performance of a vigilance task, for instance, muscle tension, as measured by electromyogram, and catecholamine levels can indicate a highly aroused state, but skin conductance might indicate that the observer is dearoused (Hovanitz, Chin, & Warm, 1989; Parasuraman, 1984). Second, it has proven difficult to define the effects of stressors on arousal independent of effects on performance (Hockey, 1986). Third, the theory can accommodate almost any results, making it a post hoc explanation that is difficult to falsify (i.e., test empirically; Hancock & Ganey, 2003; Hockey, 1984; Holland & Hancock, 1991). Finally, arousal theory assumes that a stressor (or set of stressors) affects overall processing efficiency and that differences in task demands (i.e., difficulty) are reflected only in the position of the optimal level of performance. Hockey and Hamilton

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

(1983) noted, however, that environmental stressors can have differential effects on the pattern of cognitive activity, and a single dimension, as posited by arousal theory, cannot account for such differences among stressors. Hence, a multidimensional approach is necessary in order to understand stress effects, with arousal mechanisms representing only one facet of this complex construct.

Appraisal and Regulatory Theories

Most modern theories of stress and performance have two central themes. They either explicitly include or implicitly assume an appraisal mechanism by which individuals assess their environments and select coping strategies to deal with those environments (see Hancock & Warm, 1989; Hockey, 1997; Lazarus & Folkman, 1984). Indeed, Lazarus and Folkman defined psychological stress itself as the result of an individual's appraisal of his or her environment as being taxing or exceeding his or her resources or endangering his or her well-being. The negative effects of stress are most likely to occur when individuals view an event as a threat (primary appraisal) and when they assess their coping skills as inadequate for handling the stressor (secondary appraisal; see Lazarus & Folkman, 1984). Both the person-environment interactions and the appraisal processes are likely organized at multiple levels (Matthews, 2001; Teasdale, 1999).

A second central theme of current stress theories is that individuals regulate their internal states and adapt to perturbations resulting from external stressors, including social stressors and task-based stress. Individuals respond to appraised threats (including task load) by exerting compensatory effort to either regulate their internal cognitive-affective state or to recruit resources necessary to maintain task performance. Thus, individuals are often able to maintain performance levels, particularly in real-world settings, but only at a psychological and physiological cost (Hancock & Warm, 1989; Hockey, 1997). Two current models of stress and performance that emphasize these regulatory mechanisms and adaptation are those of Hockey (1997) and Hancock and Warm (1989).

Hockey's (1997) theory is based on assumptions that behavior is goal directed and controlled by self-regulatory processes that have energetic costs associated with them. He distinguished between effort as controlled processing (e.g., working memory capacity) and effort as compensatory control (e.g., arousal level). Hockey proposed that mental resources and effort are allocated and controlled via two levels of negative feedback regulation by an effort monitor that compares current activity to goal-based performance standards. Simple, well-learned tasks are controlled at a lower level that requires very little effort (i.e., very low resource allocation) to maintain performance goals. When demands are placed on the cognitive system, via increased task demand or other forms of stress, control is exerted by a supervisory controller at the higher level. At this level, resources are recruited to compensate for goal discrepancies created by the increased demands, and information processing is more controlled and effortful. Note, however, that this represents only one potential response to stress. A second possibility would be to alter the task goals to maintain low effort or to reduce effort in the face of prolonged exposure to stress. The effortful coping response was referred to as a *strain mode*, and the reduction of effort or performance goals is *passive coping mode*. Hockey's model provides a flexible structure and an energetic mechanism by which the effects that environmental demands (i.e., stress) place on the cognitive system can be understood. As we shall see, however, this view rests on concepts that are difficult to test empirically. Specifically, the "resource concept" employed in Hockey's model, which came to dominate stress theory after the fall of the unitary arousal explanation, presents a problem in that it is ill defined and difficult to quantify. It is our contention that neuroergonomics has a significant contribution to make in improving the resource concept, since such "mental energy" must be specifically understood if neuroergonomic methods are to be effective.

The approach presented by Hancock and Warm (1989) also adopts as a fundamental tenet the idea that in many stressful situations humans adapt to their environments. This adaptation is manifested in an extended inverted-U function that describes performance change as a function of stress level, as shown in figure 13.1. Stress here can take the form of both overstimulation (hyperstress), in which the sensory systems experience an elevated level of stimulation, and understimulation (hypostress), in which the sensory systems receive disturbingly little stimulation. Note that for a wide

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

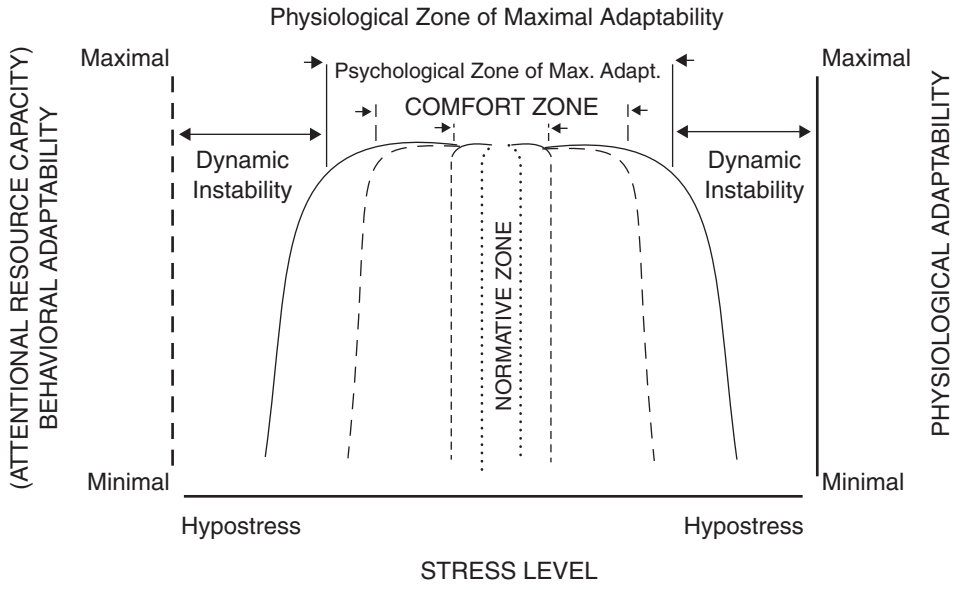


Figure 13.1. The stress-adaptation model of Hancock and Warm (1989).

range of stimulation or task demand, individuals maintain behavioral and physiological stability. There are multiple levels of adaptation that are each nested within the general, extended-U function. Thus subjective state (e.g., the normative and comfort zones in figure 13.1) is altered by relatively minor levels of disturbance, whereas it takes a greater degree of disturbance to affect behavioral performance, which is itself less robust than physiological response capacity. These different facets of response are linked: Incipient failure at one level represents the beginning of stress disturbance at the next level. While each level retains the same extended-U shape, the nesting represents the progressive fragility across levels. In the same way, there are comparable divisions within levels. For example, within the physiological level, there are progressive functions for cell, organelle, organ, and so on.

Hancock and Warm's model explicitly recognizes that tasks themselves often represent the proximal and most salient source of stress. Thus, the adaptation level of an individual will heavily depend upon the characteristics of the tasks facing that individual. Hancock and Warm (1989) proposed two fundamental dimensions along which tasks vary, these being information rate (the speed with which demands are made) and information structure (the complexity of that demand). Combined variations in task and environmental demand impose considerable stress on operators, to which they adapt via various coping efforts. Breakdown of performance under stress, and its inverse, efficient behavioral adaptability, occurs at both psychological and physiological levels, with psychological adaptability failing before comparable physiological adaptability (for a related view, see Matthews, 2001). A representation of the adaptive space in the context of the extended U is shown in figure 13.2. To locate an individual's level of adaptation to a set of stressors in an environment, one defines vectors for the level of stress and cognitive and physiological state, as well as the position of task performance along the space-time axes. Although further work is required to quantify the theoretical propositions in the Hancock and Warm (1989) model, we argue that such quantification will result in a rubric under which neuroergonomic measures of stress can be developed in coordination with performance and subjective measures. For instance, if subjective state or comfort declines prior to task performance, this should be observable not only via self-report but also using well-defined neural measures with well-validated links to cognitive processes. If the task dimensions can be specified precisely, predictions can be made regarding the level of adaptation under different task and arousal conditions. Neuroergonomic methods can thereby facilitate tests of theoretical models of stress such as that discussed here.

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

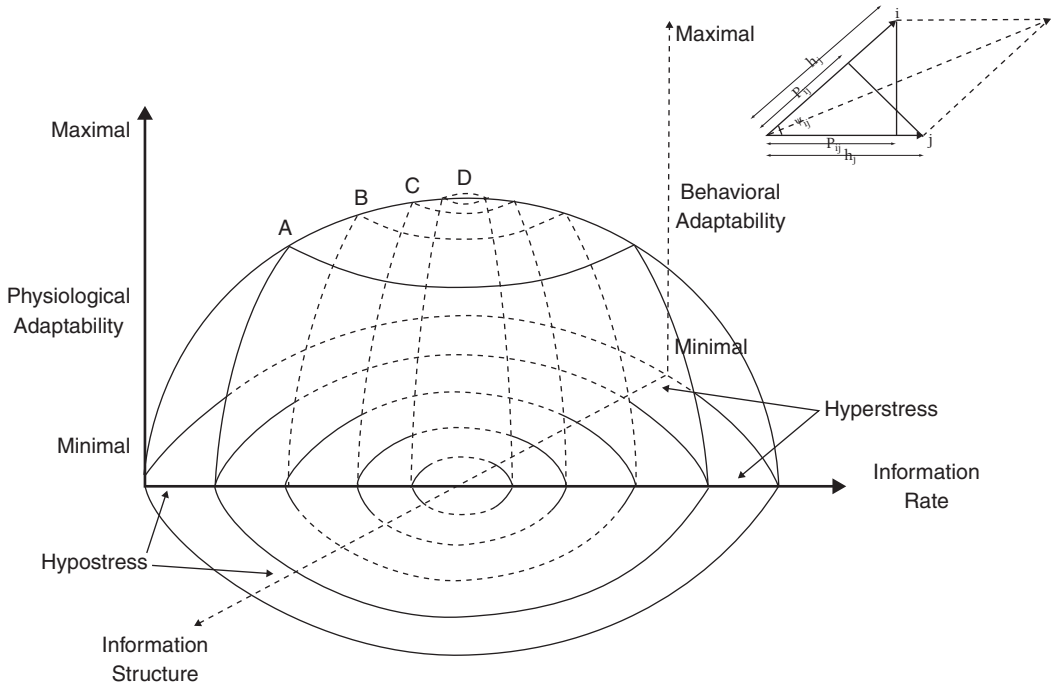


Figure 13.2. The stress-adaptation model of Hancock and Warm (1989), showing the underlying task dimensions that influence behavioral and physiological response to stress. Letters A, B, C, and D represent boundaries (tolerance limits) for physiological function (A), behavior/performance (B), subjective comfort (C), and the normative zone (D).

Stress and Neuroergonomics Research

The above considerations show that stress, like many psychological constructs, is difficult to define precisely. This represents a challenge for neuroergonomics because to identify a particular neurological state as behaviorally stressful requires a well-defined stress concept. It is also an opportunity because consideration of specific neural mechanisms for stress responses can serve to inform neuroergonomic actions at the most critical operational times. Indeed, if neuroergonomics fulfills its potential (see Hancock & Szalma, 2003; Parasuraman, 2003), it may transform the concept of stress itself. If, ultimately, cognitive states can be strongly tied to specific neurological processes, then components of stress that are currently defined psychologically (e.g., appraisal, coping, worry, distress, task engagement, etc.) may be defined in terms of their underlying neural structure and function. A note of caution is in order, however, since this reductionistic aspiration is unlikely to be fulfilled completely. While the mechanisms by which appraisals occur are likely to be universal and nomothetic, there are almost inevitably individual differences in how thorough an appraisal is, how long it takes, and what aspects of the environment are attended to (Scherer, 1999). These may not have common neurological mechanisms. Thus, there are individual differences that occur spatially (what part of the environment has drawn one's attention, and what is its relevance to the individual) and temporally (what is the time course of appraisal, how long does it take, and is more time spent appraising some criteria over others?). Neurological activity associated with appraisal mechanisms will very probably not be identical across these two dimensions. For a successful neuroergonomic approach to stress, we will need to have the capacity to examine the neural correlates of specific cognitive patterns (cf. Hockey & Hamilton, 1983) and distinguish among varieties of appraisal and coping mechanisms.

The multidimensionality of stress (Matthews, 2001) and the likely hierarchical organization of appraisal mechanisms (Teasdale, 1999) imply that neuroergonomic measures will have to be

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

sufficiently sensitive to delineate these dimensions of stress. For instance, considering the state "Big 3" (worry, distress, and task engagement; see Matthews et al., 1999, 2002), neuroergonomic tools would need to have the capacity to differentiate the neural processes underlying worry (cognitive), distress (mood, cognitive), and task engagement (mood, energetic, cognitive). Further, each of these dimensions has its specific components. For instance, task engagement consists of motivation, concentration, and energetic arousal (see Matthews et al., 2002). For a neuroergonomic system to adapt to operator stress, it will need to be sensitive to such facets of operator state. In addition, a useful neuroergonomic study of stress would be to develop valid neurological indices of the progressive "shoulders of failure" depicted in the Hancock and Warm (1989) model (see figure 13.1). For instance, indices that can track and predict the conditions under which an individual will transition from one curve to the next (e.g., from the comfort zone to failure of behavioral adaptation) would be very useful in aiding operator adaptation to stressful environments.

A core problem in development of stress theory is in the use of resource theory (Wickens, 1984) as an explanatory framework. With the fall of the unitary arousal concept, resource models emerged as a primary intervening variable to account for performance effects (see Hockey, Gaillard, & Coles, 1986). While resource theory has found some support and has been used in theoretical models of stress (e.g., Gopher, 1986; Hockey, 1997), it has been criticized as inadequate to the task of explaining attention allocation and human performance (e.g., Navon, 1984). In addition, the structure of resources and the mechanisms of resource control and allocation may not be common across individuals (Thropp, Szalma, & Hancock, 2004). The problem for stress theory is that a vague construct (resources) was employed to explain mechanisms by which another vague construct (stress) impacts information processing and performance. A significant contribution of neuroergonomics lies in its demand for precision, which only then permits computer and engineering-mediated changes in the effects of stress on information processing and performance. However, one must avoid the temptation to overly simplistic reductionism and recognize that neuroscience can enhance our understanding of resources (as both energetic states and processing capacity) but not necessarily replace psychological constructs with purely neurophysiological mechanisms. Fundamentally, the question for the neuroergonomics approach to stress is the same as for other approaches: How is it that individuals generally adapt to stress and maintain performance and what are the cognitive and perceptual mechanisms by which this occurs? The challenge for neuroergonomic efforts to answer this question will be development of neurological measures that are not merely outputs of information processing but reflect the processing itself and provide a direct measure of brain state.

To the degree that neuroergonomics can illuminate the above issues, it will also facilitate the development of more precise theory regarding the associations, dissociations, and insensitivities between performance and workload (Hancock, 1996; Oron-Gilad, Szalma, Stafford, & Hancock, 2005; Parasuraman & Hancock, 2001; Yeh & Wickens, 1988), as well as distinguishing between effortful and relatively effortless performance (cf. processing efficiency theory; see Eysenck & Calvo, 1992; and see also Hockey, 1997). Eventually, neuroergonomic approaches to stress research can lead to improving not only physiological measurement of workload but also in relating such measures to other forms of workload measurement (i.e., performance and subjective measures; see O'Donnell & Eggemeier, 1986).

Validation of Neuroergonomic Stress Measures

A fundamental problem facing those pursuing research in neuroergonomics is measurement (Hancock & Szalma, 2003). How does one connect mental processes to overt behavior (or physiological outputs) in a valid and reliable fashion? Driving this development of sound measures is the necessity for sound theory. If one adopts a monistic, reductionistic approach, then the ultimate result for neuroscience and neuroergonomics is the attempt to replace psychological models of stress with neurological models that specify the brain mechanisms that produce particular classes of appraisal and coping responses.

The alternative position that we posit here is a functionalist approach (see also Parasuraman, 2003) in which one postulates distinct physical and psychological constructs for a complete understanding

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

of stress and cognition. From this perspective, neuroscience provides another vista into understanding stress and cognition that complements psychological evidence. Whichever position is taken, one must still base neuroergonomic principles on sound theoretical models of psychological and neurological function. As we have indicated in our earlier work, understanding multidimensional concepts, including stress, requires a multimethod assessment so that a more complete picture of cognition and cognitive state can be revealed (Oron-Gilad et al., 2005).

Commonalities and Differences between Individuals

Neuroergonomics is a logical extension of the critical need for stress researchers to consider individual variation in stress response. Because stress response varies as a function of task demand and the physical, social, and organizational context, it is also likely that neuroergonomic stress profiles will vary between and within individuals. Indeed, it is already known that individuals vary in cortical arousal and limbic activation and that indices of these covary with personality traits (Eysenck, 1967; Eysenck & Eysenck, 1985), although the evidence for a causal link is mixed (e.g., see Matthews & Amelang, 1993; for a review, see Matthews, Deary, & Whiteman, 2003). Application of emerging neuroergonomic technologies to adaptive systems will provide information, combined with performance differences associated with specific traits and states, that can be used to adjust systems to particular operators and to adapt as operator states change over time. Further, application of neuroscience techniques might enhance our theoretical understating of individual differences in stress response. Although many theories of stress have been developed, comprehensive theory on individual differences in stress and coping in ergonomic domains has been lacking. Neuroergonomics offers a new set of tools for individual differences researchers for both empirical investigation and theory development.

Individual differences in cognition, personality, and other aspects of behavior have traditionally been examined using the psychometric approach. Understanding the sources of such differences has also been illuminated by behavioral genetic studies of psychometric test performance. For example, this method has been used to show that general intelligence, or *g,* is highly heritable (Plomin & Crabbe, 2000). However, conventional behavioral genetics cannot identify the particular genes involved in intelligence or personality. The spectacular advances in molecular genetics now allow a complementary approach to behavioral genetics—*allelic association*. In this approach, normal variations in single genes, identified using DNA genotyping, are associated with individual differences in performance on cognitive tests. This method has been applied to the study of individual differences in cognition in healthy individuals, revealing evidence of modulation of attention and working memory by specific genes (Parasuraman, Greenwood, Kumar, & Fossella, 2005). Parasuraman and Caggiano (2005) have incorporated this approach into their neuroergonomic framework and discussed how molecular genetics can pinpoint the sources of individual differences and thereby provide new insights into traditional issues of selection and training. This approach could also be applied to examining individual differences in stress response.

Hedonomics and Positive Psychology

Thus far we have been concerned primarily with implications of neuroergonomics for stress. There is also an opportunity to explore the application of neuroergonomics to the antithesis of stressful conditions. The latter study has been termed *hedonomics*, which is defined as "that branch of science which facilitates the pleasant or enjoyable aspects of human-technology interaction" (Hancock, Pepe, & Murphy, 2005, p. 8). This represents an effort not simply to alleviate the bad but also to promote the good.

Studies of the level of pleasure experienced by individuals are rooted in classic effects of limbic stimulation on behavior (Olds & Milner, 1954). Indeed, neuroergonomic indices of attention might clarify the processes that occur for maladaptive versus adaptive attentional narrowing, recovery from stress and performance degradation (i.e., hysteresis). In addition, neuroergonomics can contribute to the emerging positive psychology trend (Seligman & Csikszentmihalyi, 2000) by identifying the neurological processes underlying flow states (Csikszentmihalyi, 1990) in which individuals are fully engaged in a task and information processing is

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

more automatic and fluid rather than controlled and effortful.

Ethical Issues in Neuroergonomics and Stress

There is a pervasive question that must attend the development of all new human-machine technologies, and that is the issue of purpose (Hancock, 1997). Subsumed under purpose are issues such as morality, benefit, ethics, aesthetics, cost, and the like. As we have merely introduced elsewhere (Hancock & Szalma, 2003), there are issues of privacy, information ownership, and freedom that will also have to be considered as cognitive neuroscience develops and is applied to the design of work. In regard to stress, there is a particular danger that those who are naturally prone to specific patterns of stress response (e.g., those high in neuroticism or trait anxiety) may be excluded from opportunities or even punished by controlling authorities for their personality characteristics. If we assume, however, that the rights of individuals have been secured, application of neuroergonomic technologies to monitor emotional states of individuals could provide useful information for adjusting system status to the characteristics of the individual. Thus, those who are prone to trait anxiety could be assisted by an automated system when it is detected that their state anxiety is increasing to levels that make errors or other performance failures more likely. In such an application, the technology could serve to increase the performance of anxious individuals to a level comparable to that of individuals low in trait anxiety. More generally, neuroergonomics offers a new avenue for research and application in individual differences in coping and responses to stress, as a technology of inclusion rather than a technology of exclusion. In many modern systems, the environment is sufficiently flexible that it can be adapted to the characteristics and current cognitive state of the individual operator. This can improve performance, reduce the workload and stress associated with the task, and perhaps even render the task more enjoyable.

As neuroergonomic interventions emerge, consideration of intention and privacy will be crucial. In the age of ever-increasing technology, the opportunity for any individual to preserve his or her own private experience is vastly diminished. The proliferation of video technology alone means that events that have previously remained hidden behind a barrier of institutional silence now become open to public inspection. The modern threat of terrorism has also engendered new categories of surveillance technologies which use sophisticated software that seeks to process the nominal intent of individuals. Added to these developments are the evolutions in detection technologies that use evidence such as DNA traces to track the presence of specific individuals. Such developments indicate that overt actions are now detectable and recordable and thus potentially punishable. Further, this invasion is moving to the realm of speech and communication. Recorded by sophisticated technologies, utterances regarding (allegedly) maleficent intent now become culpable evidence. Despite a nominal freedom of expression, one cannot offer threats, conspire to harm, or engage in certain forms of conversation without the threat of detection and punishment. Many would argue that destructive acts should be punished, and that voicing violent intent is also culpable, whether the intent is fulfilled or not (hence the ambivalence regarding jokes concerning exploding devices at airport screening facilities). However, neuroergonomics, if the vision is fully or even partly fulfilled, now promises to extend these trends further. It will not only be actions and language that could be considered culpable but, more dangerously, the thought itself. As Marcus Aurelius rightly noted, we are the sole observer and arbiter of our own personal private experience. We may safely and unimpeachably think unpleasant thoughts about anyone or anything, content in the knowledge that our thoughts are private and impenetrable to any agency or individual. Neuroergonomics could threaten this fundamental right and indeed, in so doing, threaten what it is to be an individual human being.

We paint this picture not to discourage the pursuit of neuroergonomics, but rather to sound a cautionary alarm. While neuroergonomic interventions have the potential to provide early warning that an individual is overstressed and therefore could be used to mitigate the negative effects of such stress, the imposition of the neuroergonomic intervention could itself impose significant stress on the individual, particularly if those in authority (e.g., company management, governmental agencies) have access to the information provided by neurological and physiological measures, and individuals appraise

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

this situation as a potential threat with which they cannot effectively cope.

Conclusion

Clearly, one application of neuroergonomics to stress will be a set of additional tools to monitor operator state for markers predictive of later performance failure and, based on such information, adjusting system behavior to mitigate the stress. However, to reduce stress effectively requires an understanding of its etiology and the mechanisms by which it occurs and affects performance. To address these issues requires development of good theoretical models, and we see the neuroergonomic approach to stress research making a significant contribution toward this development. As a multidimensional construct, stress requires multidimensional assessment, and improvement of theory requires that the relations among the dimensions be well articulated. Neuroergonomics will facilitate this process. The key will be to develop reliable and valid neurological metrics relating to cognitive and energetic states and the processes underlying the recruitment, allocation, and depletion of cognitive resources. Our optimism is not unguarded, however. First, neurological measures should be viewed as one piece of a complex puzzle rather than as a reductive replacement for other indices of stress. Omission of psychological constructs and measures would weaken the positive impact of neuroscience for stress research and stress mitigation efforts. Second, in implementing stress research, we must ensure that the application of neuroergonomics to stress mitigation does not increase stress by imposing violations of security and privacy for the individual. If these concerns are adequately addressed, neuroergonomics could not only revolutionize stress as a psychological construct but could also serve to transform the experience of stress itself.

On a more general front, neuroergonomics could be used to mitigate stress completely. With sufficient understanding of neural stress states and their precursors in both the environment and the operator's appraisal of that environment, it would be possible, in theory, to develop a servomechanistic system whereby sources of stress were ablated as soon as they arose. But would this be a good thing? It is not that we enjoy certain adverse situations, but it may be that it is the stimulation of such adverse conditions that spur us to higher achievement. This brings us to our final observations on the very thorny issue of the purpose of technology itself. Is the purpose of technology to eradicate all human need and as a corollary to this, to instantly and effortlessly grant all human physical desires? We suggest not. Indeed, such a state of apparent *dolce far niente* (life without care) might well prove the equivalent of the medieval view of hell! Further, it is currently unclear how to respond to mistakes in neuroergonomics if the purpose is to facilitate the immediate transition from an intention to an action. Complex error recovery processes are built into the current human motor system—will such effective guards be embedded into future neuroergonomics systems? Some scientists have opined that all technology is value neutral in that it can be employed for both laudable and detestable purposes. But this is a flawed argument because the creation of each new embodiment of technology is an intentional act that itself expresses an intrinsic value-based decision (see Hancock, 1997). To whatever degree that value is apparently hidden in the process of conception, design, and fabrication, and to whatever degree others choose to pervert that original intention, the act itself implies value. Thus, we need at the present stage of development to consider not whether we can develop neuroergonomics technologies but rather, whether we should. Needless to say, this is liable to be a somewhat futile discussion since rarely, if ever, in human history have we refrained from doing what is conceived as being possible, whether we should or not. Indeed, it is this very motivation that may well be the demise of the species. To end on a more hopeful note: perhaps not. Perhaps the present capitalist-driven global structure will consider the greater good of all individuals (and indeed all life) and refrain from the crass, materialistic exploitation of whatever innovations are realized. Then again—perhaps not. And that was a hopeful note.

Main Points

1. Stress is a multidimensional construct that requires multidimensional assessment. Neuroergonomics promises to provide valuable tools for this effort.

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

2. A major problem for stress research is the difficulty in precisely defining the concepts of stress and cognitive resources. Neuroergonomic efforts should be directed toward elucidating these constructs.
3. Neuroergonomics can improve the state of stress theory via programmatic research toward establishing the links between neurological and cognitive states. This is critical, since the validity of neuroergonomic measures depends heavily on sound psychological theory.
4. A potential practical application for stress mitigation will be the ability to monitor operator state in real time so that systems can adapt to those states as operators experience stress. Such efforts, in the context of neuroergonomics, are already underway.
5. Neuroergonomics can also be useful for studying individual differences in stress and coping and establishing general theoretical framework for individual differences in performance, stress, and workload.
6. While the promise for neuroergonomics is high, we must ensure that individuals' privacy and well-being are preserved so that the cure does not become worse than the disease.

Key Readings

Hancock, P. A., & Desmond, P. A. (Eds.). (2001). *Stress, workload, and fatigue*. Mahwah, NJ: Erlbaum.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519–537.

Hockey, G. R. J., Gaillard, A. W. K., & Coles, M. G. H. (Eds.). (1986). *Energetics and human information processing*. Dordrecht: Martinus Nijhoff.

Hockey, R., & Hamilton, P. (1983). The cognitive patterning of stress states. In: G. R. J. Hockey (Ed.), *Stress and fatigue in human performance* (pp. 331–362). Chichester: Wiley.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer-Verlag.

References

Asterita, M. F. (1985). *The physiology of stress*. New York: Human Sciences Press.

Cannon, W. (1932). *The wisdom of the body*. New York: Norton.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper.

Dunbar, F. (1954). *Emotion and bodily changes*. New York: Columbia University Press.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*, 183–201.

Elliot, G. R., & Eisdorfer, C. (1982). *Stress and human health*. New York: Springer.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Charles Thomas.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. New York: Plenum.

Eysenck, M. W., & Calvo, M. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion*, *6*, 409–434.

Frankenhaeuser, M. (1986). A psychobiological framework for research on human stress and coping. In M. H. Appley & R. Trumball (Eds.), *Dynamics of stress: Physiological, psychological, and social perspectives* (pp. 101–116). New York: Plenum.

Gopher, D. (1986). In defence of resources: On structure, energies, pools, and the allocation of attention. In G. R. J. Hockey, A.W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 353–371). Dordrecht: Martinus Nijhoff.

Hancock, P. A. (1996). Effects of control order, augmented feedback, input device and practice on tracking performance and perceived workload. *Ergonomics*, *39*, 1146–1162.

Hancock, P. A. (1997). *Essays on the future of human-machine systems*. Eden Prairie, MN: Banta.

Hancock, P. A., & Desmond, P. A. (Eds.). (2001). *Stress, workload, and fatigue*. Mahwah, NJ: Erlbaum.

Hancock, P. A., & Dirkin, G. R. (1983). Stressor induced attentional narrowing: Implications for design and operation of person-machine systems. *Proceedings of the Human Factors Association of Canada, 16*, 19–21.

Hancock, P. A., & Ganey, H. C. N. (2003). From the inverted-U to the extended-U: The evolution of a law of psychology. *Human Performance in Extreme Environments*, 7(1), 5–14.

Hancock, P. A., Pepe, A., & Murphy, L. L. (2005). Hedonomics: The power of positive and pleasurable ergonomics. *Ergonomics in Design*, *13*, 8–14.

Hancock, P. A., & Szalma, J. L. (2003). The future of neuroergonomics. *Theoretical Issues in Ergonomics Science, 4*, 238–249.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519–537.

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

Hancock, P. A., & Weaver, J. L. (2005). On time distortion under stress. *Theoretical Issues in Ergonomics Science, 6,* 193–211.

Hebb, D. O. (1955). Drives and the CNS (conceptual nervous system). *Psychological Review, 62*, 243–254.

Hockey, G. R. J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of human perception and performance: Vol. II. Cognitive processes and performance* (pp. 1–49). New York: Wiley.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*, 73–93.

Hockey, G. R. J., Gaillard, A. W. K., & Coles, M. G. H. (Eds.). (1986). *Energetics and human information processing*. Dordrecht: Martinus Nijhoff.

Hockey, R. (1984). Varieties of attentional state: The effects of environment. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 449–483). New York: Academic Press.

Hockey, R., & Hamilton, P. (1983). The cognitive patterning of stress states. In G. R. J. Hockey (Ed.), *Stress and fatigue in human performance* (pp. 331–362). Chichester: Wiley.

Holland, F. G., & Hancock, P. A. (1991, June). *The inverted-U: A paradigm in chaos*. Paper presented at the annual meeting of the North American Society for the Psychology of Sport and Physical Activity, Asilomar, CA.

Hovanitz, C. A., Chin, K., & Warm, J. S. (1989). Complexities in life stress-dysfunction relationships: A case in point—tension headache. *Journal of Behavioral Medicine, 12*, 55–75.

Jones, D. (1984). Performance effects. In D. M. Jones & A. J. Chapman (Eds.), *Noise and society* (pp. 155–184) Chichester: Wiley.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer-Verlag.

Matthews, G. (2001). Levels of transaction: A cognitive sciences framework for operator stress. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 5–33). Mahwah, NJ: Erlbaum.

Matthews, G., & Amelang, M. (1993). Extraversion, arousal theory, and performance: A study of individual differences in the EEG. *Personality and Individual Differences, 14*, 347–364.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, J. A., Huggins, J., Gilliland, K., et al. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion, 2*, 315–340.

Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits* (2nd ed.). Cambridge, UK: Cambridge University Press.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three"? In I. Mervielde, I. J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 335–350). Tilburg: Tilburg University Press.

McGrath, J. J. (1970). A conceptual framework for research on stress. In J. J. McGrath (Ed.), *Social and psychological factors in stress* (pp. 10–21). New York: Holt, Rinehart, and Winston.

Navon, D. (1984). Resources: A theoretical soup stone? *Psychological Review, 91*, 216–234.

O'Donnell, R. D. & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of human performance: Vol. 2. Cognitive processes and performance* (pp. 1–49). New York: Wiley.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions in the rat brain. *Journal of Comparative and Physiological Psychology, 49*, 281–285.

Oron-Gilad, T., Szalma, J. L., Stafford, S. C., & Hancock, P. A. (2005). *On the relationship between workload and performance*. Manuscript submitted for publication.

Parasuraman, R. (1984). The psychobiology of sustained attention. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 61–101). Chichester: Wiley.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science, 1–2*, 5–20.

Parasuraman, R., & Caggiano, D. (2005). Neural and genetic assays of mental workload. In D. McBride & D. Schmorrow (Eds.), *Quantifying human information processing* (pp. 123–155). Lanham, MD: Rowman and Littlefield.

Parasuraman, R., Greenwood, P. M., Kumar, R., & Fossella, J. (2005). Beyond heritability: Neurotransmitter genes differentially modulate visuospatial attention and working memory. *Psychological Science*, *16*(3), 200–207.

Parasuraman, R., & Hancock, P. A. (2001). Adaptive control of mental workload. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 305–320). Mahwah, NJ: Erlbaum.

Pilcher, J. J., Nadler, E., & Busch, C. (2002). Effects of hot and cold temperature exposure on performance: A meta-analytic review. *Ergonomics, 45*, 682–698.

Plomin, R., & Crabbe, J. (2000). DNA. *Psychological Bulletin, 126*, 806–828.

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

Scherer, K. R. (1999). Appraisal theory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 637–663). Chichester: Wiley.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5–14.

Selye, H. (1976). *The stress of life* (Rev. ed.). New York: McGraw-Hill.

Szalma, J. L., & Hancock, P. A. (2002). *On mental resources and performance under stress*. White Paper, MIT^2 Laboratory, University of Central Florida. Available at www.mit.ucf.edu.

Teasdale, J. D. (1999). Multi-level theories of cognition-emotion relations. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 665–681). Chichester: Wiley.

Thropp, J. E., Szalma, J. L., & Hancock, P. A. (2004). Performance operating characteristics for spatial and temporal discriminations: Common or separate capacities? *Proceedings of the Human Factors and Ergonomics Society, 48,* 1880–1884.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.) *Varieties of attention* (pp. 63–102). San Diego: Academic Press.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Yeh, Y. Y., & Wickens, C. D. (1988). Dissociations of performance and subjective measures of workload. *Human Factors*, *30*, 111–120.

Hancock, P.A. & Szalma, J.L. (2006). Stress and Neuroergonomics. In: R. Parasuraman and M. Rizzo (Eds.), Neuroergonomics: The brain at work. Oxford: Oxford University Press (pp 195-206)

A META-ANALYSIS OF APERIODIC NOISE STRESS ON HUMAN PERFORMANCE

B.M. Saxton*, J.M. Ross, A. Braczyk, G.E. Conway, J.L. Szalma, and P.A. Hancock
University of Central Florida
Orlando, Florida, 32816

ABSTRACT

Aperiodic noise, also known as intermittent noise, is a pervasive and influential source of stress across military environments, and can be defined by the changes in its intensity over a given period of time (therefore containing 'gaps' between louder phases of the noise). With examples ranging from the discharge of weapons to vehicle and machinery movements, then it is intuitive to recognize that this common form of noise may constitute a risk to Soldier performance across a range of tasks (i.e., as measured by speed and accuracy metrics). In order to quantify these effects, a meta-analytic evaluation of aperiodic noise effects on performance was undertaken. The results indicate that a general effect of aperiodic noise is to exert a negative influence on performance; however this effect is contingent upon the type of tasks and performance measures used. These results can be used to inform decisions concerning when noise should be mitigated or even alternatively exploited in military settings.

1. INTRODUCTION

Humans are constantly exposed to environmental stressors, one of which is noise stress. Noise stress includes, "*all disturbances, annoying and hazardous sounds influencing the hearing organ and other senses of the human body"* (Engel, Augustnska, Koton, & Kacmarska, 2006, pp 1826). Though many researchers have examined the relationship between noise stress and human performance, the results have been mixed regarding the direction and size of the effect. In order to potentially summarize this confusing situation, a quantitative meta-analysis was preformed to obtain a global effect-size to determine whether noise stress facilitates or debilitates human performance.

2. METHODS

A sampling of the noise stress and human performance literature was conducted and 425 reports were acquired for the meta-analysis. The articles were obtained through literary search engines using PsychINFO, MEDLINE, and Dissertation Abstracts International Databases. Using these literary search engines the primary search terms "noise" and "speech" were combined with the following secondary search terms: "memory", "decision making", "problem solving", "attention", "vigilance", "tracking", "marksmanship"/"shooting", "fine motor", and "gross motor."

In order to ensure meaningful combination of results each of the research papers was evaluated based on five inclusion criteria. The first criterion was that the paper had to contain an empirical study examining intermittent noise stress as an independent variable. The second criterion was that the experiment must compare a control (lower dB) and experimental (higher dB) conditions. Third, each study must contain a dependent variable that measured human performance. Fourth, the sample within the experiment must be healthy adult participants, and finally, each study had to present enough information to obtain effect size statistics. Out of the original 425 acquired articles, 107 separate effect sizes were calculated, which resulted in a 25% inclusion rate.

Performance types were categorized using the taxonomy used in the US Army's Improved Performance Research Integration Tool (IMPRINT) performance modeling tool. Task types were initially separated into in four higher-level categories, according to main type of demand made on the human performer by the task at hand. Performance was therefore categorized as perceptual, cognitive, motor, or communication tasks. These task types were then further divided into nine lower level categories, in order to recognize the differential demands within each of the higher order categories. The nine categories were therefore: perceptual (visual recognition/discrimination), cognitive (numerical analysis), cognitive (information processing/problem solving), motor (fine motor discrete), motor (fine motor continuous), motor (gross motor light), motor (gross motor heavy), communication (reading & writing), and communication (oral). Examples of these categories are as given below, to aid clarity.

Perceptual (visual recognition/discrimination) tasks are those that require a visual identification of a target

(e.g., seeing a moving target in the distance). Cognitive (numerical analysis) tasks are those that require the performer to carry out mental arithmetic calculations while cognitive (information processing/problem solving) tasks are those that require the mental processing of information for the purposes of forming a final conclusion (e.g., deciding the best firing location). Motor (fine motor discrete) tasks require little effort in performing distinct motor actions within a particular sequence (i.e. pressing a series of buttons), whereas a fine motor continuous task requires a participant to maintain an uninterrupted series of movement actions within a certain level of performance (e.g.. tracking task using a joystick or driving a vehicle). Gross motor (heavy) tasks require a large physical effort to perform the required action (e.g., moving a large object). Gross motor (light) tasks require movements featuring multiple body parts, though with little physical effort (e.g., cleaning a weapon). A communication (reading & writing) task requires reading and/writing text (e.g., a checklist), and communication (oral) tasks require the participant to listen or talk to another person (e.g., receiving an order).

Separate categorizations were made according to the type of performance measure used, in terms of whether the tasks emphasized *speed* or *accuracy* of the response. As meta analyses allow a hierarchical approach, it was also possible to assess the degree to which effects on one level (e.g., the task type) were moderated by effects on another level (e.g., whether speed or accuracy was emphasized as a response criterion). Though these hierarchical analyses allow the opportunity to examine specific combinations of factors, it must be recognized that the more specific the data are, the less data there is. Hence, some moderator analyses contain a limited or no number of studies.

The data were analyzed using calculations of effect size. As few papers provided effect sizes in their results, effects were computed using several equations, depending on the data provided. Equation 1 was used to calculate effect sizes when means and standard deviations were reported.

$$g = (X_e - X_c)/SD_c \text{----------(1)}$$

Within the formula X_E = the mean of the experimental condition whereas X_C = the mean of the control condition. For within-subjects SD_C = standard deviation for the control condition but in between-subjects design SD_{pooled} = the pooled standard deviation of the experimental and control condition, which was obtained using Equation 2. The data received from Equation 2 was then applied in Equation 3 to get a between-subjects pooled effect size.

$$S_{pooled} = SQRT(((\underline{n_E}\text{-}1)(\underline{SD_E^2})+(\underline{n_C}\text{-}1)(\underline{SD_C^2})) / (n_e+n_c-2)) \text{----------(2)}$$

$$g = (X_e - X_c)/SD_{pooled} \text{---------- (3)}$$

Equation 4 was used to calculate effect sizes when an F-test with one degree of freedom was presented in a between-subjects article, while Equation 5 was used in a within-subjects design.

$$g = (SQRT(F\ test))*(((1/n_e)+(1/n_c))^(1/2)) \text{-------- (4)}$$

$$g = (SQRT(F\ test))\ /\ (SQRT(n)) \text{----------(5)}$$

Equation 6 was used to convert t-test value(s) into F-test value(s), which were then used in calculating effect sizes.

$$t\ test = (SQRT(F\ test)) \text{----------(6)}$$

3 RESULTS

Hedges' *g,* a standardized mean difference between the experimental and control conditions (Hedges & Olkin, 1985) was used as a measure of effect-size, which allows effects from different tasks and measures to be synthesized. The direction of effect-size was standardized so that positive effect-sizes represent performance improvements in the experimental group relative to the control, whereas negative effect-sizes reflect performance impairment in the experimental condition. Cohen's (1988) guidelines were used to provide an estimation of effect size, with 0.2 representing a small effect, 0.5 a medium effect and 0.8 a large effect. In the current analysis, we collapsed over common participant samples at each level of the analysis to avoid violating the assumed independence between effect-size estimates. In addition to the weighted mean effect size, two variance estimates were computed: variability due to sampling error (s_e^2) and variability of effect sizes (s_g^2). A large s_δ^2 indicates there is variability among the observed effect sizes that cannot be accounted for by the sampling error and one or more variables moderating the magnitude of the effect is in question.

The 'global' (overall) effect of aperiodic noise on performance was drawn from 107 total qualifying studies and resulted in a moderate negative effect (g = -0.53), indicating that aperiodic noise exerts a degrading effect on performance. A secondary analysis examined performance on the different kinds of task, specifically perceptual, cognitive, psychomotor, and communication tasks. Aperiodic noise was found to have a moderate negative effect on cognitive performance (g = -0.51; see Table 1). The effects of noise stress on communicative performance was found to result in a moderate negative effect (g =

-0.45) while a small negative effect (g = -0.36) was found for noise stress effects on motor performance. A large positive effect (g = 0.63) was found when aperiodic noise was applied during perceptual performance. It must be highlighted that the results for perception, motor and communication tasks were drawn from a limited number of available studies, and so the data must be interpreted with caution. Differential effects of aperiodic noise were found when the type of dependent measure used (i.e., speed or accuracy) was analyzed. Noise was found to exert a large negative effect on performance accuracy (g = -0.81), whilst a moderate positive effect (g = 0.58) was found for performance speed (see Table 2).

A hierarchical analysis was conducted to determine whether the effects of aperiodic noise on the task types were moderated by the type of metric used within the tasks (i.e., whether the measures were of speed or accuracy, see Table 3). While under noise stress, cognitive tasks emphasizing performance accuracy were found to suffer large negative effects (g = -0.84), whereas tasks emphasizing performance speed were found to be facilitated under noise (a moderate positive effect, g = 0.58).

Unfortunately, a low number of contributing studies for the remaining combinations of dependent measure and task type hindered our ability to obtain meaningful results.

Table 1. *Moderator Analysis of Task Category*

Intermittent	**k**	**average g**	**s_g^2**	**s_e^2**	**s_d^2**	**s_e**	**LL 95% CI**	**UL CI**	**95%**
perception	3	0.63	0.28	0.82	0	0.52	-0.4	1.65	
cognitive	95	-0.51	1.04	0.63	0.41	0.08	-0.67	-0.35	
motor	3	-0.36	0.23	0.05	0.17	13	-0.61	-0.1	
communication	7	-0.45	0.7	0.78	0	0.33	-1.11	0.2	

Table 2. *Moderator Analysis of Dependent Measure*

Intermittent	**k**	**average g**	**s_g^2**	**s_e^2**	**s_d^2**	**s_e**	**LL 95% CI**	**UL CI**	**95%**
accuracy	103	-0.81	0.67	0.62	0.06	0.08	-0.96	-0.66	
speed	17	0.58	11.77	0.28	11.5	0.13	0.33	0.83	

Table 3. *Moderator Analysis of Task Category within each Dependent Measure Category*

Intermittent	**k**	**average g**	**s_g^2**	**s_e^2**	**s_d^2**	**s_e**	**LL 95% CI**	**UL CI**	**95%**
Accuracy									
perception	2	0.58							
cognitive	93	-0.84	0.78	0.68	0.1	0.09	-1.01	-0.66	
motor	2	-0.64							
communication	7	-0.45	0.7	0.78	0	0.33	-1.11	0.2	
Speed									
perception	2	1.06							
cognitive	14	0.58	11.89	0.23	11.65	0.13	0.32	0.83	
motor	1	0.49							
communication	0								

4. DISCUSSION

The results indicate that aperiodic noise generally exerts a degrading effect on performance. Unfortunately, the majority of the studies examined only cognitive performance, limiting the ability to analyze the other task categories. This highlights the need for additional research in these areas in order to 'fill in the gaps' in knowledge. Cognitive tasks were found to degrade under noise, though differential effects were found when the type of measure used was considered. Cognitive tasks that required performance accuracy was found to degrade under noise, whilst cognitive tasks that required fast responding was facilitated under noise. These findings have important implications for Soldiers engaging in any form of task in which information processing is a key component. With the increasing use of technology in the modern army (such as wearable computers and an increasing emphasis on netcentric warfare), the disruption of Soldier cognitive performance by environmental noise constitutes a significant threat to mission success and Soldier safety.

5. CONCLUSIONS

Though the overall effect of aperiodic noise is debilitative to goal-directed performance, our analysis demonstrates that this is not always the case. Rather, there are certain combinations of task/environmental conditions under which aperiodic noise actually acts to facilitate performance (e.g., perception tasks requiring accurate responses). Aperiodic noise is a common stressor in military environments, with every Soldier exposed at some point. Accordingly, it is important to understand the differential effects of noise on the range of task environments, thereby increasing the chances for task or mission success. Tasks vulnerable to debilitative noise effects can be recognized and steps can be taken to mitigate the threat, and tasks that are aided under noise can similarly be recognized, with a view to maximizing the potential for performance facilitation.

ACKNOWLEDGMENTS

The information contain within this paper is based on the information gained throughout the analysis and viewpoints of the authors. The Department of the Army, Department of Defense, or any Unites States government affiliation should be attributed with the views presented within the paper. The research was conducted on a sub-contract provided by Micro-Analysis and Design (MAAD) from the Army Research Laboratory (ARL).

REFERENCES

Cohen, J., (1988): Statistical Power Analysis for the Behavioral Sciences, Hilldale, NJ: Lawerence Erlbaum Associates.

Engel, Z., Augustnska, D., Koton, J., and Kacmarska, A., 2006: Noise: Definitions. In W. Karwowski *International Encyclopedia of Ergonomics and Human Factors*, pp1826

Hedges, L.V., and Olkin, J.A., (1985): Statistcal Methods for Meta-Analysis, New York: Academic Press.

Theologus, G.C., Wheaton, G.R., and Fleishman, E.A, 1974: Effect of Intermittent, Moderate Intensity Noise Stress on Human Performance, *Journal of Applied Psychology*, 59, 539-547.

COMMUNICATING THROUGH THE USE OF VIBROTACTILE DISPLAYS FOR DISMOUNTED AND MOUNTED SOLDIERS

James L. Merlo
Peter I. Terrence
Shawn Stafford
Richard Gilson, Ph.D.
P.A Hancock, Ph.D.
University of Central Florida

Elizabeth S. Redden, Ph.D.
Andrea Krausman
Christian B. Carstens, Ph.D
Rodger Pettitt
Timothy L. White
U. S. Army Research Laboratory/ HRED

ABSTRACT

The purpose of the studies reported here was to determine if participants wearing a purposed-designed tactile display could accurately report cue localization and messaging while undergoing different levels of physiological stress. Experiment 1 found that participants could effectively receive tactile messaging while navigating a physically challenging obstacle course. Experiment 2 demonstrated that tactile localization could be achieved by participants experiencing extreme whole body vibration produced by a vehicle simulator which replicated movement of different military vehicles in extreme conditions. Collective results illustrate that the current tactile display has significant potential for communication and/or directional cueing in demanding, real-world conditions.

INTRODUCTION

Combat conditions can impose significant demands on Soldier senses, limiting their ability to communicate through normal auditory and visual pathways. Noisy (e.g., weapon fire, vehicle engines) and murky (e.g., smoke, sandstorm) conditions can hinder the ability to communicate critical data, such as relevant threat information or simple squad movement instructions. In an environment replete with visual and auditory signals and noise, one way to circumvent this issue is to communicate through a relatively unused information channel: touch. Cutaneous communication systems have enjoyed considerable success in many domains, such as aircraft stick shakers, cellular telephone vibratory alerts, and reading using Braille. However, tactile displays are not as pervasive in military settings and may offer a relatively unexploited sensory channel for soldier communications. The goals of the present studies were to assess the performance of wearable tactile informational display systems in controlled field conditions for both dismounted and mounted operations and evaluate their relevance to systems like the Army Future Combat Systems and Future Force Warrior.

According to multiple resource theory (see Wickens, 1984, 2002), parsing information across the input modalities can alleviate sensory bottlenecks in conditions of high workload, and reduce interference across visual and auditory channels. Tactile displays offer additional advantages. First, they are non-illuminating and acoustically covert, allowing the soldier maintain a stealth advantage. Conversely, traditional visual and auditory displays can mask important environmental information, such as distant enemy movement or approaching footsteps. Tactile displays also offer the advantages of self referent-presence and omni-directionality.

Our research group, in partnership with industry, has designed a wearable tactile display capable of delivering patterns of vibratory stimulation at multiple loci. This system is able to convey information clearly beyond simple alerts or directional cueing. The display allows for precise control of frequency, gain, and onset times. With this level of stimulus control, consistent patterns can communicate complex messages, as well as simple alerts. Stimulus parameters have been derived, based upon feedback from a group of subject matter experts consisting of retired US Soldiers and Marines, to tactually convey key hand and arm signals. The reason for this was three-fold. First, hand and arm signal movements have a spatio-temporal patterns that can be conveyed via comparable patterns applied to the skin. These tactile signals could be presented covertly, without soldier movements that may signal their position to an enemy force. Second, the wireless transmission of these signals to the tactile display increases the likelihood that all squad members will receive a signal simultaneously. For example, when a team leader is informed of a potential threat and visually signals a “Halt” command, the soldiers in front of the team leader may not visually acquire the command while scanning their surroundings and maintaining local security. This can lead to an “accordion” effect, where by soldiers do not immediately respond to halt and become spread out. Third, Soldiers are already well trained on hand and arm signals. Therefore, it is reasonable to expect a degree of transfer of training to learn the tactual form, provided that the tactile patterns are designed intuitively to closely approximate their known visual counterparts. A tactile lexicon was developed using commonly used commands of the existing U.S. Army hand and arm signals (identified by the SME group; Department of the Army, 1987) allowing eye-free and ear-free communication between display wearers (see Gilson, Redden, & Elliott, in press). Concurrent with laboratory research, there was the need to determine if such a tactile displays could

function in applied settings involving significant physical and cognitive demands.

The Mechanism of Display

For tactile displays that are based on vibration, the key skin receptors are the Pacinian corpuscles, which consist of nerve endings surrounded concentrically by layers of non-neural connective tissue. Pacinian corpuscles respond most readily to vibration at frequencies around 200-300 Hz, whereas, the free nerve endings are sensitive to much lower frequencies between 50-100 Hz (Bear, Connors, & Paradiso, 2001). Researchers have had problems when attempting to operationalize tactile displays since difficulties arise in the detection of the signal. For example, Sklar and Sarter (1999) experienced challenges with tactile signals being readily identified on the wrist and arm. Gilland and Schlegel (1994) reported that tactile communication applied to the head lowered performance on a concurrent task. These, and similar issues, present significant challenges to those seeking to use tactile displays in the real world.

As vibrotactile stimulator placement on the body has had the aforementioned difficulties on the head and extremities, the torso appears to be the preferred placement site for a wireless, fieldable system. The torso offers the least opportunity for tactors to shift during demanding physical tasks and also (based on SME input) would be the least likely location for it to interfere with combat tasks and for the stimuli to remain easily perceptible.

Our laboratory research has demonstrated that tactile cueing yields significantly faster and more accurate performance across varying lab test than comparable spatial auditory cues. Further research has demonstrated this finding is relatively stable across a variety of body orientations, even when spatial translation is required (Terrence, Brill, & Gilson, 2005) and under physiological stress (Merlo, Stafford, Gilson, & Hancock, 2006). The overall 99.4% accuracy rate displayed by the participants in the latter, physiological stress study was highly encouraging in respect to the potential of the current tactile display design. The accuracy of the messages and the reported intuitiveness with which they were received was also a testament to the potential utility of the present 'language' transformation format.

The natural progression for the present tactile system has moved beyond simple directional cueing and into the realm of more advanced covert communication. The purpose of the studies described here was to ensure that the tactile display can be useful in situations that are physiologically stressful and in a more applied setting than just running on a treadmill. Tactile signals that are only recognizable in pristine, quiet conditions may have limited use in real-world military operations. These studies show the results of the tactile systems as they were tested in two separate operational experiments, one for dismounted and one for mounted soldiers.

Materials and Apparatus

The tactile system used in the present studies, the TACTile Information Communication System (TACTICS), was developed by researchers using a plunger-type vibrotactile actuator (hereafter referred to as a "tactor"), whereas much of the previously published research have used systems incorporating a 'pancake type' tactor, commonly used in cellular telephones. To determine if tactile communication and signaling would be viable under physiological stress, a durable field-worthy system was created using the plunger-type tactor due to its frequency and concentrated stimulus characteristics. The model C2 tactors, manufactured by Engineering Acoustics, Inc. (see Figure 1) are essentially acoustic transducers that displace 200-300 Hz sinusoidal vibrations onto the skin. This frequency range, in combination with their 17 gram mass, is sufficient to activate the skin's Pacinian corpuscles. The C2's contactor is 7 mm, with a 1 mm gap separating it from the tactor aluminum housing. The C2 is a tuned device, meaning it operates well only within a very restricted frequency range, in this case the optimal frequency of 250 Hz. These devices offer several benefits over pancake-motor type tactors that rely on a spinning arm to vibrate the entire tactor housing. First, the C2 tactor housing remains stationary as only the center plunger oscillates to provide a concentrated vibratory stimulus to a specific point on the skin. Second, pancake-motor tactors do not allow for independent control of frequency and amplitude. In order to raise the frequency to the range most readily perceived by the Pacinian corpuscle, the amplitude most also increase because the motor must spin faster and therefore increasingly displace the entire housing.



Figure 1. *An enlargement of a single tactor, model C2, manufactured by Engineering Acoustics, Inc.*

The number of tactors used on the different parts of the body should carefully be considered. Cholewiak, Brill, and Schwab (2004) found that for the torso, a ring of eight loci of vibration was the most efficient configuration with identification accuracy exceeding 90%.

The system uses a self-contained, wearable control box that provides power and signal generation for the tactors (see Figure 2). The control box houses a Bluetooth wireless communication device so that stimuli can be activated remotely from a Pocket PC or laptop computer.

The tactors themselves, as well as their associated wiring, are embedded in an elasticized belt worn around the torso slightly above navel height. The eight tactors in the system are spaced equidistantly around the abdomen with tactor 1 placed just above the navel and tactor 5 on the spine. The equidistant configuration is maintained due to the elasticity of the belt, allowing for ease of use from one user to the next without reconfiguring the display. Thus, there is a tactor every 45 degrees from the navel all the way around the body, (45 degrees x 8 tactors = 360 degrees).

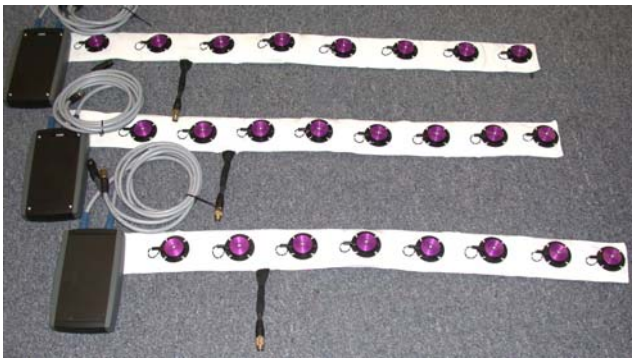


Figure 2. *Three tactile displays belt assemblies are shown above along with their Tactor Controller Boxes (TCB). Each box includes a wireless Bluetooth receiver and the controller circuitry and waveform generator.*

The experiments presented here were conducted to assess the utility of complex tactile communication in both a mounted and dismounted environment. The first experiment assesses performance during individual movement techniques through an obstacle course. The second assessed performance in a vehicle simulator that was also providing whole body vibration consistent with military vehicle motion over selected terrains.

EXPERIMENT 1

Experimental Participants

For this experiment, twenty-nine male Soldiers from the Infantry Training Brigade (ITB) located at Fort Benning, Georgia participated in the assessment of the tactile messaging system on an obstacle course. The mean age of the participants was 21 years and they were all in self-reported good physical health and they all could perform the assigned tasks of negotiating the obstacles.

Equipment

For the experiment, the tactile system presented four exemplar tactile patterns designed to be analogous to standard Army hand and arm signals (Pettitt, Redden, & Carstens, 2006). Tactile messages were created using the standard Army hand and arm signals (Department of the Army, 1987) as a design guideline. The four signals chosen for the experiment were “Halt”, “Rally”, “Move Out”, and “NBC (nuclear, biological, chemical warning).” The tactile representations of these signals were designed in a collaborative effort between the present researchers and the aforementioned group of SMEs.

Experimental Procedure

Soldiers completed an obstacle course while receiving communication from either the tactile system or from a leader to their front or rear who provided visual hand and arm signals.

Soldiers were instructed to perform urban and non-urban tactical maneuvers and to conduct what is termed Individual Movement Techniques (IMT). The obstacles were categorized according to the activity the Soldier was performing at the time he was given a command. The four obstacle categorizations were patrolling, crawling, firing and climbing. Soldiers initially walked through the course, where each obstacle and position was explained before they began the first recorded trial. All Soldiers also completed one familiarization trial wearing their uniform, standard fighting load, tactile belt, and carrying their assigned weapon. The tactile signals that were used in practice were not the same as the signals used in the testing. The times and locations that the Soldier received the different signals were consistent among blocks. However, the individual Soldier never knew when or where they would receive either a hand or tactile signal.

Questionnaires were administered to elicit Soldiers’ opinions about their performance and experiences while wearing the tactile system. The questionnaires were designed to enable Soldiers to rate the ease of receiving and understanding visual and tactile communications while negotiating obstacles, as well as their overall experience with the tactile system. Questionnaires were administered to each Soldier upon the completion of each experimental block on the course.

Each soldier was given approximately ten minutes of familiarization and training with the tactile system before beginning the obstacle course. Familiarization for every Soldier consisted of approximately 17 repetitions of the four tactile signals: three repetitions without wearing fighting loads, 4 times while wearing fighting loads, and 2 times each while wearing fighting loads in the kneeling, prone, combat roll, walk, and run positions/actions.

Soldiers were also presented with a refresher course on visual hand and arm signals. Soldiers were trained until they were 100 % accurate on all signals. Each Soldier was retested on all of the signals before each obstacle course experimental block. Upon completion of the training, the Soldiers were given a questionnaire designed to assess their perception of the training adequacy. Before the experimental trials were conducted, Soldiers participated in a practice trial on the obstacle course during which hand and arm signals and tactile signals were sent.

Soldiers completed three experimental blocks on the obstacle course. During each block, individual Soldiers wore the tactile system and were led through the obstacle course as if in a tactical scenario led by their team leader and followed by a squad leader. These positions are consistent with a Soldier acting as a member of a squad in a wedge formation**.** The team leader was designated to communicate visual hand and arm signals from the front and the squad leader from the rear. Tactile commands were communicated to the Soldier by a controller operating the control unit within blue tooth range of the squad member. A data collector recorded the time from the command's initiation until it was acknowledged by the Soldier and whether the correct response was given. As a visual distraction task, Soldiers were told to look for Special Forces, Airborne and Ranger tabs (military uniform insignia) that were placed at various locations on the course and to scan the woods for targets. At the completion of each block, a questionnaire was administered.

Results

Only a short period of time was required to train the Soldiers to become accurate on interpreting the tactile signals and for refresher training on the hand and arm signals. Soldiers became proficient after approximately ten minutes of individual training on the tactile signals and rated them as easier to learn than the hand and arm signals. All the Soldiers rated the training as being "good" to "extremely good".

Due to a floor effect in the time to respond to signals (the lowest possible time is 0.1 sec), the data are highly skewed. Analysis of variance (ANOVA) is based on the assumption that the data are approximately normally distributed. Thus all ANOVAs and follow-on comparisons reported here were completed using the log transformed data. However, when the log transformed data and the untransformed data were analyzed, the results were essentially identical. Table 1 shows the mean response times across the three IMT iterations for the three methods of signaling. For purposes of analysis, failures to respond to signals were coded as response times of 20 sec.

Table 1. *Mean Response Times in seconds, Obstacle Course Trials*

Signal Modality	**Mean**	***SD***
Hand signals -- rear	4.65	2.02
Hand signals -- front	2.93	1.17
Tactile signals	1.81	0.64

Table 2. *Mean Log Response Times in seconds, Obstacle Course Trials*

Signal Modality	**Mean**	***SD***
Hand signals -- rear	0.6325	0.1730
Hand signals -- front	0.4411	0.1429
Tactile signals	0.2393	0.1229

A repeated-measures ANOVA was performed on the log transformed scores (see Table 2). This analysis yielded a statistically significant effect for signal modality, $F(2, 58) = 80.3$, $p < .001$, $\eta_p^2 = .735$. Follow-up pairwise comparisons were done using Holm's Sequential Bonferroni correction to control for family-wise error (see Table 3). These follow-on comparisons show that response times were significantly faster with the tactile signals than with the hand signals, and that response times were significantly faster when the hand signals came from the Soldier's front rather than the rear.

Table 3. *Follow-on Comparisons Mean Log Response Times, Obstacle Course Trials*

Comparison	***df***	***t***	**Obtained *p***	**Required *p***
Rear vs. Front	29	5.92	< .001*	0.05
Rear vs. Tactile	29	12.03	< .001*	0.0167
Front vs. Tactile	29	7.25	< .001*	0.025

**p < .05, two-tailed*

Soldiers rated both the "hand signals" and the "tactile signals" as each being very easy to learn. Soldiers rated both the "tactile commands" and the "front hand commands" as very easy to detect and interpret. One Soldier stated that the tactile system seemed to become progressively easier to interpret as the iterations progressed. By the third iteration, he stated that the tactile belt was easier and quicker to understand than clearly visible hand and arm signals. The "rear hand signals" were rated as being more difficult to detect and interpret.

An additional analysis on missed responses was conducted to categorize them into "failures to detect" the commands versus "incorrect responses" to detected commands. A Chi Square analysis was conducted on this data and showed there was a significant difference in the number of "failures to detect" among the three signal

modalities: $\chi^2(2, N = 30) = 18.5, p < .01$. There was also a significant difference in the number of "failures to detect" among the four obstacle types: $\chi^2(3, N = 30) = 7.82, p < .05$. The greatest number of "failures to detect" came with the "hand signals from the rear" and on the "climbing obstacles". There was a non-significant trend for the greatest number of incorrect responses occurring with the "hand signals from the rear" and "on the climbing obstacles".

Discussion

The tactile signal patterns were found to be intuitive and easy for the Soldiers to understand. Very little training time (less than 10 minutes) was required for Soldiers to become accurate on interpreting the four tactile signals used during the experiment. Results demonstrated Soldiers performing an obstacle course were able to receive, interpret and accurately respond to the tactile commands faster than when the information was passed by a leader in the front of a wedge formation or by a leader in the back of a wedge formation using conventional hand-and-arm signals. Soldiers also commented they were better able to focus more attention on negotiating obstacles and on local area SA when receiving tactile signals than when maintaining visual contact with their leaders in order to receive standard hand and arm signals.

There is reason to expect that tactile signals would be even more effective in more realistic operational context. The obstacle course made it much easier to see the hand and arm signals than it would have been if the experiment had been conducted for example in a wooded environment where vegetation and terrain features would have masked the leaders. The obstacle course made conditions identical for each run and therefore more amenable to anticipation thought consistency. Also the experiment was conducted only during daylight conditions which again made it easier to see the hand and arm signals than it would have been if the experiment had been conducted at night.. Soldiers indicated the tactile system allowed them to focus more attention on negotiating obstacles, and that it would be useful in tactical situations in which they would need to focus on other tasks such as security. Soldiers commented that it was very difficult to receive a visual hand and arm signal at certain points on the course where their full attention was given to negotiating the obstacle, or when they could not maintain visual contact with the leaders. Soldiers stated they knew immediately when they received a tactile signal no matter what obstacle they were negotiating.

The use of a tactile communication system has the potential to improve infantry team performance beyond that which is documented here. During this experiment leaders in the front and rear of the Soldiers were not obscured by terrain, vegetation or darkness. In other words, the conditions of this experiment were optimal for the Soldiers' abilities to see the conventional hand and arm signals. During combat situations, larger dispersions and obscurants could greatly inhibit reception of visual hand and arm signals. Visual barriers in an urban combat situation could impair hand and arm signaling. Also, hand and arm signals are traditionally passed along throughout the squad so that the time that the first squad member receives the signal could be much quicker than the time that the signal is passed to and received by the last squad member.

A tactile communication system allows simultaneous reception of signals by all squad members. For example, a "halt" signal sent by visual signals could result in a wave effect so that the last squad member to receive the signal could still be moving long well the time when the squad needed to stop. A "halt" signal sent by a tactile system can be received by all squad members in less than two seconds. A further benefit provided by a tactile system is the increased local situational awareness (SA) experienced by the squad because the tactile system would free their eyes from having to watch for visual signals (Smith & Hancock, 1995). A third benefit of adding a tactile system is that Soldiers would have two means of receiving communication because the visual hand and arm signals would still be available.

EXPERIMENT 2

The objective of this experiment was to determine how vehicle vibration affects detection and localization of vibrotactile signals.

Experimental Participants

Twelve individuals from the pool of civilian personnel at Aberdeen Proving Ground (APG) voluntarily participated in this study. There were 9 males and 3 females with ages ranging from 24 to 54 years. The average age was 37.7.

Equipment

A six-degree-of-freedom Ride Motion Simulator (RMS) was used to simulate movement of a High Mobility Multipurpose Wheeled Vehicle (HMMWV) and a Bradley Fighting Vehicle (BFV) moving over gravel and cross country terrain at an average speed of 20 mph (see Figure 3). Tactile signals were sent via the aforementioned tactile display. The C2 tactor was driven at two signal strengths. In the first condition, identified as TACTICS 1, the tactor signal frequency was set to the standard 250 Hz and at the optimal intensity used in previous experimentation. In the second condition, identified as TACTICS 2, the signal strength was analogous to the characteristics of a "pancake motor" tactor that vibrates at around 60 Hz. Stimulus matching was conducted in the laboratory until appropriate gain

settings subjectively matched the gain of the C2 operating at 250 Hz to the subjective operating intensity of the "pancake" motor-type tactor. Each signal consisted of one tactor vibrating for 0.5 seconds. The two conditions allowed not only for exploration in the effects of vehicular movement on tactile displays but also a comparison of two different tactile presentations, in this case a subjectively matched low frequency tactor to the tactile display used in the aforementioned studies.



Figure 3. *A six-degree-of-freedom Ride Motion Simulator (RMS) was used to simulate movement of a High Mobility Multipurpose Wheeled Vehicle (HMMWV) and a Bradley Fighting Vehicle (BFV) moving over gravel and cross country terrain at an average speed of 20 mph.*

Experimental Procedure

Each participant wore the tactile display around the abdomen, situated just above the navel. The experimental conditions were counterbalanced and subjective data about the systems were collected at the end of the experiment. The primary task of the participants was to detect and verbally state the spatial location of the tactile stimuli. Participants completed two training sessions, one with each tactile system. For each training session, 16 tactile signals were presented and participants verbally indicated the location of the vibration they received using the eight locations of the tactors in accordance with compass directions. Upon completion of the training, each participant was seated on the ride motion simulator and completed a practice trial to become familiar with how the simulator operated. Participants also received tactile signals at different locations around their waist to become familiar with how these signals felt. Again, participants verbally indicated the location of each signal.

For the experiment, participants were seated on the stationary RMS and baseline measures for each tactile system were taken. For the baseline, participants sat on the stationary RMS and received 16 tactile signals (two at each of the eight compass points), verbally indicating the location of the vibration they received (northwest, south, etc.). After the baseline measures were completed, participants were assigned a treatment condition and began the ride portion of the experiment. During the simulated ride, the tactile system generated signals at irregular intervals. Participants verbally indicated the location of the vibration, and no verbal report was coded as failure to perceive the stimulus. Response time was measured and each treatment condition lasted for approximately 3 minutes. Participants received a total of 16 tactile signals, two at each of the eight locations. Interstimulus intervals were randomized to reduce participant anticipation of the signals.

Results

Data were analyzed according to the percentage of detected signals, the percentage of correctly localized signals, and response time. In the baseline, participants were able to detect 100% of the tactile signals. Response time was equivalent for both systems. For the moving trials, 98.4% of the TACTICS 1 signals were detected and 93.8% of the TACTICS 2 signals were detected. Of those signals detected, 97% of the signals were correctly localized for both systems. Response time for both systems was 1.6 seconds. With respect to errors, the majority were localizations that were off by 45 degrees. Of the eight locations on the display, "West", was the most frequently missed. It was originally hypothesized that "South" would be the most frequently missed due to its location over the indentation in the lower back, however, the data did not support this contention.

DISCUSSION

Clearly, tactile signals are registered and useful, even during high physiological stress and significant arousal situations. In contrast to some other prototype systems, the present one does not use the pager or "pancake like" motors that have been used in past research. The piston type tactor design and our present placement strategy appear to be very effective in maximizing signal reception. The potential for such tactile communications and the use of tactile displays for enhancing soldier performance seems particularly opportune at the present time. It has also not escaped notice that this system has a great potential for use by other military services and personnel in the civilian sector who also seek to increase communication capacity and alleviate visual and auditory overload in demanding operational environments.

ACKNOWLEDGEMENTS

This work was supported by ARL under the ATO for Situational Understanding with special thanks to Dr. Elizabeth Redden, ARL, HRED and to the Department of Defense Multidisciplinary University Research Initiative (MURI) program, P.A. Hancock, Principal Investigator, administered by the Army Research Office under Grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy.

REFERENCES

Bear, M.F., Connors, B.W., & Paradiso, M.A. (2001). *Neuroscience: Exploring the Brain*. 2nd Edition. Lippincott Williams and Wilkins; Baltimore, MD.

Cholewiak, R.W., Brill, J.C., & Schwab, A. (2004). Vibrotactile localization on the abdomen: Effects of place and space. *Perception & Psychophysics, 66*(6), 970-987.

Department of the Army. (1987). *Visual signals*. (Field Manual No. 21-60). Washington, DC: Government Printing Office.

Gilson, R.D., Redden, E.S., Elliott, L.R. (in press) *Remote tactile displays for future soldiers.* Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Merlo, J.L., Stafford S.C., Gilson, R. & Hancock, P.A. (2006). The effects of physiological stress on tactile communication. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, CA.

Pettitt, R.A., Redden, E.S., & Carstens, C.C. (2006). *A Comparison of Hand and Arm Signals to a Covert Tactile Communication System in a Dynamic Environment*. ARL-TR-3838. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Sklar, A.E., & Sarter, N.B. (1999). Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. *Human Factors, 41,* 543–552.

Smith, K., & Hancock, P.A. (1995). Situation awareness is adaptive, externally-directed consciousness. *Human Factors, 37 (1)*, 137-148.

Terrence, P.I., Brill, J.C., & Gilson, R.D. (2005). Body orientation and the perception of spatial auditory and tactile cues. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, Orlando, FL.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), *Varieties of attention.* (pp. 63-101). Hillsdale, NJ: Erlbaum.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science, 3*(2), 159-177.

A case study of information and communication technology in emergency management training

Rogier Woltjer*

Cognitive Systems Engineering Laboratory
Department of Computer and Information Science
Linköpings Universitet, SE-581 83 Linköping, Sweden
E-mail: rogwo@ida.liu.se
*Corresponding author

Ida Lindgren and Kip Smith

Division of Industrial Ergonomics
Department of Mechanical Engineering
Linköpings Universitet, SE-581 83 Linköping, Sweden
E-mail: idali@ikp.liu.se
E-mail: kipsm@ikp.liu.se

Abstract: This paper addresses the roles of Information and Communication Technology (ICT) in training for effective emergency management and inter-organisational coordination. Collocation can encourage the development of common ground and trust and, in turn, result in greater efficiency and effectiveness. We expect to find communication and artefact use during collocated training that cannot readily transfer to the ICT used to link distributed work settings. This expectation makes the reliance on ICT and distributed work during emergency management operations suspect. To test these claims, we observed a large-scale, real-time exercise designed to facilitate cooperation among electricity and telecommunications companies. The exercise scenario was similar to the January 2005 windstorm that left much of southern Sweden without electricity or telephone service and revealed the need for better cooperation among utility providers. The observations suggest that while collocation is clearly beneficial, a mismatch in ICT use between collocated training and distributed emergency management operations is likely to be detrimental for preparedness.

Keywords: emergency management; training; artefact use; collocated work; distributed work; real-time exercise; coordination; cooperation.

Reference to this paper should be made as follows: Woltjer, R., Lindgren, I. and Smith, K. (xxxx) 'A case study of information and communication technology in emergency management training', *Int. J. Emergency Management*, Vol. X, No. Y, pp.000–000.

Biographical notes: Rogier Woltjer is a doctoral candidate in the field of Cognitive Systems Engineering at Linköpings Universitet, Sweden. His research focuses on the analysis of performance, training, and decision support in emergency and incident management and command and control. He received his MSc from Vrije Universiteit Amsterdam and his LicEng degree from Linköpings Universitet.

Copyright © 200x Inderscience Enterprises Ltd.

Ida Lindgren is a doctoral student of Industrial Ergonomics at Linköpings Universitet, Sweden. Her research concerns bridging cultural barriers to collaborative decision-making in emergency relief teams. Her research interests also include social psychology, emergency management, training, and communication. She received her MSc in Cognitive Science from Linköpings Universitet.

Kip Smith is Professor of Industrial Ergonomics at Linköpings Universitet. His research interests include collaborative decision-making, the impact of distributed command and control on trust and performance, and neuroeconomics. He received his PhD from the Carlson School of Management at the University of Minnesota.

1 Introduction

Emergency management requires effective coordination. Sound training can overcome the persistent problem of ineffective coordination. This paper addresses the role of Information and Communication Technology (ICT) in training for effective emergency management and inter-organisational coordination. The literature on work organisation suggests that groups that work in a collocated manner (*i.e.*, face-to-face in the same physical location), rather than in a distributed manner, benefit from better communication and more flexible interaction and task allocation. These benefits of collocation facilitate the development of common ground and trust and, in turn, result in greater efficiency and effectiveness. We expect to find similar benefits of collocation in emergency management operations. Further, we expect to find communication and artefact use during collocated training that cannot readily transfer to the ICT used to link distributed work settings. These expectations do not bode well for the reliance on ICT and distributed work during emergency management operations. To test these claims, we observed a large-scale, real-time exercise designed to facilitate cooperation among electricity and telecommunications companies. The exercise scenario was similar to the January 2005 windstorm that left much of southern Sweden without electricity or telephone service and revealed the need for better cooperation among utility providers. The observations suggest that while collocation is clearly beneficial, a mismatch in ICT use between collocated training and distributed emergency management operations is likely to be detrimental for preparedness.

2 Background

Emergency management has been defined as “the process of coordinating an emergency or its aftermath by communication with participants and organising the deployment and use of emergency resources” (Alexander, 2003, p.118). Emergency management is a dynamic process conducted under stressful conditions that demands flexible but rigorous planning, cooperation among stakeholders, heedfulness, vigilance, and follow-through. The premise of this article is that ICT plays multiple roles that may facilitate the management of and training for emergency operations designed to restore power and telecommunications in response to catastrophic breakdowns in service.

In this article, the acronym ICT refers to any technical system or artefact designed to gather, process, distribute or mediate information. We consider two disparate types of work organisation – collocated work (*i.e.*, in the same physical location) and distributed work. Distributed work requires artefact-mediated communication, whereas collocated work affords face-to-face communication. This difference influences artefact use and utility. Among the artefacts we consider are telephones, whiteboards and web-based tools. We use the word 'utilities' to refer to telecommunications companies, electric utilities and related service providers that develop and maintain the power and telecommunications infrastructures.

This article begins with an analysis of the Swedish power grid and telecommunications networks and their characteristics. These considerations support the development of claims about the impacts of ICT and collocated work on emergency management of the power grid. We then discuss training, collocated work and a training exercise conducted in Sweden designed to meet the need for cooperation and preparedness among the utilities. We describe the exercise, its players and the roles played by ICT during the exercise. The article concludes with implications of the observational study for training and the use of ICT in emergency management and its training. It is our hope that these findings may be applied to any domain in which incident management, training and recovery from failure of complex systems is an issue.

3 The problem

Ineffective coordination is a reoccurring problem in emergency management (McEntire, 2002). It is often exacerbated by commercial rivalry or mutual distrust between stakeholders. The devastating storm that reached the west coast of Sweden on the 8th of January 2005 showed the need for Swedish utilities to learn how to cooperate and to train for emergency management and recovery operations. Hurricane-force winds ripped countless trees out of the ground or broke them like matchsticks. Cellular phone masts were blown away. The wind and fallen timber disabled virtually all power and telecommunications in rural southern Sweden. Approximately 415 000 households were left without electricity or telephone service or both (Alexandersson, 2005). In the aftermath of the storm, the utilities were unable to work together to formulate and implement a speedy response. It took several weeks for the utilities and their subcontractors to return all service to normal.

3.1 The power and telecom networks as interdependent systems

The power grid is a tightly coupled system (Perrow, 1984) that consists of many power lines and transfer stations that connect a relatively small number of large, established companies to each other and their customers. There are few redundancies (*e.g.*, alternative power lines/stations) and few possibilities for 'storage' (*e.g.*, batteries) designed into the system. The interactions between components are so complex that it can be difficult to identify and isolate failed components. Cascading failures of system components can create a domino effect that crosses company boundaries (Barron, 2003; Brown, 2004) and forces commercial rivals to come together to solve their joint problem. Swedish electric utilities have acknowledged their tight coupling and complex interactions and have responded to emergencies by working together to repair the shared

infrastructure. Prior to the January 2005 storm, they had formed 'electricity coordination teams' responsible for organising their preparedness for an extensive power failure and for coordinating shared resources during emergencies. In spite of this effort, they had great difficulty managing the power failure following the storm.

In contrast, the recent proliferation of mobile telephones and telecom providers has created a highly divisive environment with little incentive to cooperate. Each operator taps into a network characterised by fairly linear (non-complex) interactions. Failed system components can be identified relatively easily and are easily substituted. As a result, telecom companies appear to have adopted a stand-alone attitude. Until recently, many were openly dismissive of efforts to work together to prepare for the day when their tightly coupled shared infrastructure breaks down. Cooperation in the aftermath of the storm was minimal which appears to have slowed the recovery process. Following the lead of the electric utilities, Swedish telecom utilities are in the process of forming a group responsible for inter-company cooperation during and after emergencies.

A crisis like the January 2005 storm reveals the interdependencies between the electric and telecommunications systems. These two types of utilities are loosely coupled (Orton and Weick, 1990) to each other. Telecom equipment runs on electricity and is, accordingly, dependent on the power grid. Conversely, recovery from a breakdown of the electricity network relies mainly on functioning telecommunications. When an emergency occurs, all are affected and a joint effort is needed for quick recovery.

3.2 The message and the question

After the storm, Swedish society delivered a clear message to the utilities – cooperation and preparedness is necessary. Private companies and government agencies that supply electricity and telecommunications were put on alert that they must educate and train their personnel to be prepared to quickly repair breakdowns in service like those caused by the storm. An open question is, accordingly, how best to train those responsible for managing the utilities' efforts to recover from an emergency.

3.3 Training

Military forces and civil services continuously train to stand prepared when an emergency hits a country or region. They are expected to be able to handle any foreseeable crisis (*e.g.*, a natural disaster or terrorist attack). It is unfortunate that similar training and preparedness have not been the norm for utilities and corporations responsible for maintaining electric and telecommunications networks. Since the electricity and telecom utilities are loosely coupled and a joint effort is needed for quick recovery, training for preparedness should emphasise joint efforts and interdependencies of the networks.

Training allows the emergency management staff to develop their personal relationships and is likely to increase emergency response effectiveness (Baldwin, 1994; Perry and Lindell, 2003). Unfortunately, management often balks when it is asked to commit personnel and resources to emergency training exercises. Officials often fail to see the benefit (return on investment) of recurrent training (Perry and Lindell, 2003; Regester and Larkin, 2005). However, highly realistic scenarios (regarding scale, workload, artefacts, pace, procedures, *etc.*) can be instrumental in convincing these officials of the benefits associated with training (Regester and Larkin, 2005).

Highly realistic exercises train both domain-specific skills (*e.g.*, use of artefacts and ICT) and domain-general skills (*e.g.*, cooperation and communication). People inevitably learn new facts, skills and procedures when they are active in a training environment (Säljö, 2000). Training environments therefore should situate the trainees in an appropriate setting that facilitates learning the skills and knowledge relevant to their actual work environment. This consideration extends to the ICT systems present in training and actual work settings. If there is a mismatch between the (ICT systems in the) training environment and the world of real practice, the domain-specific skills acquired during training may not transfer. Alternatively, acquired skills may transfer but be inappropriate and become sources of confusion and ineffectiveness. Moreover, emergency management systems not used on a regular basis are unlikely to be of use in actual emergencies (Turoff, 2002).

3.4 Collocated work

Teams that train face-to-face and work in separate locations may encounter a mismatch between the artefacts used in the training environment and those used in actual practice. This mismatch is a cause for real concern because communication styles and artefact use learned during collocated training may be inappropriate when working in physically dispersed locations.

When a team is working in a collocated manner, the members share the same physical location (Olson and Olson, 2000). There are distinct advantages to collocated work compared to distributed work. Collocation can be beneficial for:

- communication (Olson and Olson, 2000; Olson *et al.*, 2002)
- flexible organisation (Olson *et al.*, 2002)
- the development of trust (McAllister, 1995)
- the development of common ground (Clark, 1996)
- the working environment and productivity (Olson *et al.*, 2002).

However, Jarvenpaa and Leidner (1999) found that 'swift trust' (Meyerson *et al.*, 1996) can develop for a short period of time even in virtual teams, in which members have never met in person.

The opposite of collocated work – distributed work – is usually chosen when the task at hand is short term in nature and/or it is undesirable to relocate members of the team from their workplaces. Members of distributed teams typically communicate using telecom media such as telephone, e-mail and videoconferencing (Cramton, 2002).

In spite of the advantages of collocated work, sharing a common space is not suitable for all kinds of tasks or people (Schunn *et al.*, 2002). There are two main disadvantages:

1 the situation can sometimes seem chaotic when people are coming and going and the personnel can overhear each other's conversations

2 constantly having people around leaves no room for privacy.

However, there is considerable evidence that individuals working in a distributed manner are likely to benefit from meeting face-to-face once in a while (Hinds and Kiesler, 2002).

4 ICT in emergency management

An important contrast between collocated and distributed settings is the difference in the artefacts that are available. Distributed teams typically rely on ICT to coordinate their activities. Some ICT used in distributed settings enable people to communicate in a manner that emulates face-to-face communication (Olson and Olson, 2000). The use and design of these digital artefacts is therefore of interest in modern emergency management.

In order to understand how people think, use artefacts, learn and master situations, it must be acknowledged that work is accomplished by a combination of people, the artefacts they use and the organisation they are part of (Hollnagel and Woods, 2005; Säljö, 2000). In this article, the artefacts of interest include but are not limited to ICT artefacts. We adopt Hollnagel and Woods' (2005, p.66) definition of artefact: "an artefact is a device used to carry out or facilitate a specific function that is part of work". Through the use of ICT, people can solve problems and control situations that they could not address by themselves. In order for artefacts to serve people's purposes, they:

1 should be as transparent as possible so that artefact and human become a joint system with a common purpose

2 should amplify human skills, rather than replace human activity and reduce their skills (Hollnagel and Woods, 2005).

It is important to note that when learning how to use an artefact, it must not be taken out of the context in which it is to be used. If artefact design meets the above recommendations, they can make us smart (Hutchins, 1995; Norman, 1993; Säljö, 2000).

Observations in many domains imply that there are always benefits and disadvantages to the use of ICT, and this is also true for emergency management (Suparamaniam, 2003). Technology enables coordination, but the efficacy of that coordination is dependent on the correct functioning of the technology. ICT can:

1 enable communication between physically dispersed people

2 connect people in the field to their home organisations and thereby provide confidence and security

3 enable people to send large amounts of data in short periods of time.

But if ICT is not designed or functioning properly, the meaning of data and information can be lost and effective coordination jeopardised (Suparamaniam, 2003). Another issue with ICT in general and in emergency management specifically is that it can be difficult for the receiver of information to verify the source, validity and accuracy of the information, *i.e.*, to ascertain whether the information can be trusted. In sum, ICT can facilitate effective coordination, but only when it works and only when the receiver of the information can verify its reliability and understand its content.

Claims

These considerations support the following three claims.

1 Considering the advantages of collocated work, the team assembled for the utilities' exercise is likely to benefit from being able to communicate face-to-face and from the opportunity to develop trust and common ground about the task of emergency management coordination.

2 A second benefit of collocated training is the opportunity to refer to the same artefacts while communicating face-to-face and for the team to develop patterns of interaction with those artefacts that may transfer to distributed work and, hence, enhance operations.

3 However, considering the potential for a mismatch in communication, task allocation and artefact (ICT) use during collocated training and actual distributed work, the associated skills and knowledge gained during training may not readily transfer to distributed work during actual emergency management operations.

In short, while there are clear benefits of collocated training for distributed work, these benefits may be reduced by a mismatch in ICT use.

5 Method

This section discusses the large-scale, real-time exercise conducted by representatives of the Swedish electricity and telecom utilities and other critical infrastructure. The immediate goal of the exercise was to develop cooperation and communication between decision-makers from various electricity and telecommunication companies in situations of crisis. The long-term goal was to develop knowledge of the practice of crisis management and its consequences for society.

5.1 The exercise, its scenario and the work environment

The exercise was performed in real-time and simulated the effects of a devastating storm. The scenario combined hypothetical events with events similar to the January 2005 storm's effects on power and communication infrastructures. Parts of the exercise had been planned long before the January 2005 storm. The exercise was initially meant to train radio correspondents' crisis communication skills. The repercussions of the January storm increased interest in the exercise. Because the scenario was only partially based on an actual emergency, the participants were likely to experience it as highly realistic without being overly constrained by their experience with the actual emergency (Alexander, 2000).

The scenario was played in real-time at a training centre and in parallel at various remote sites on 19 April 2005. The exercise took approximately ten hours. It started with the situation that on 14 April, a big storm forcefully accelerated through the Benelux, Denmark and southern Norway through the southwest of Sweden towards Stockholm on the 15th. The violent advance of the storm resulted in power outages and downed telephone lines and masts, and associated severe infrastructure problems. The utilities became interested in participating to be able to re-enact the consequences of the storm and learn to deal with them more efficiently.

5.2 Participants

Electricity and telecommunications utilities, as well as many other infrastructure organisations and companies, sent a total of approximately 60 representatives to the exercise. Participants were decision-makers from most of the companies involved in the real storm who were responsible for reacting to the emergency and for coordinating

activities to return service to normal. We observed a team of seven persons responsible for coordinating cooperation between electric utilities during emergency operations. These seven are the actual members of one of the official Electricity Coordination Teams. The team consisted of six men and one woman, of whom five were representatives of electric companies and two of governmental agencies. Because the equivalent telecommunications coordination team was meeting for the first time, they did not want to be observed.

The team we observed was assembled by the electric companies and agencies and was the same team that would be responding to a real-life crisis. One of the team members is the designated chairperson. All other team members have flexible roles. The participants have extensive domain knowledge in their area of expertise. They normally work in different locations and communicate by telecom links. During the exercise, they were collocated and therefore used a different set of artefacts than they were used to. Informed consent for video-taping, anonymity, publication, *etc.*, was obtained.

5.3 Task

The team's task was to function as a coordination team, facilitating the cooperation between the companies. To do this, the team had to:

- collect information
- assess the situation before and during the power failure
- distribute material, equipment, personnel and other resources from other parts of the country to the affected area
- maintain continuous contact with the agency responsible for the power grid
- find the people and competence needed to cope with the given situation.

The Swedish electric companies have formulated guidelines for implementing inter-organisational cooperation and coordination. The starting point for these guidelines is that the individual companies have the responsibility to make sure that there are resources (*i.e.*, material, equipment, personnel and competence) to cover any potential power failure and that the resources are organised in a way that minimises the impact of a power failure on society. To be able to do this, a company representative should be reachable through telecommunications at all times and have access to a Power Grid Cooperation System (PGCS). The PGCS is an internet-based computer system displaying the status of the power grid. The system updates in real time with data about calls for production and repair work, customers, power grid sections, companies, *etc.*

Many of the team's tasks are conducted with the help of the PGCS. This system is a tool for planning and supporting emergency management during power failures. The electric utilities' special interest group developed the system and is responsible for its maintenance and for making it available to electric companies around the country. The system is internet-based and therefore assumes that electricity is available when using it. The emergency management group must therefore have access to some power source to be able to use the PGCS. The team uses the PGCS to register, assess and distribute information about what is happening in the affected area. They can also use the system to request actions and resources from companies within the area. All decisions the team makes and notes from their meetings must be made available in the PGCS for others to see.

The team uses a second system – a Crisis Information Message System (CIMS) – to disseminate information to the media and other external parties during an emergency. The CIMS is an internet-based system administered by the Swedish Radio network. It serves the purpose of swiftly spreading information about the emergency to receivers of the electric companies' choice. Releasing information to the public is assumed to decrease the pressure on the utilities' customer service desks. The CIMS was not used during the exercise.

5.4 Data collection

We observed the team in its workrooms. Data collection methods included:

- direct observation by two observers
- video recording of the team's workplace
- traces of artefact use (telephone, white-board, internet)
- a transcript of a semistructured retrospective interview with a subset of the observed participants.

5.5 Exercise procedure

There were two groups of people involved in the exercise. The first group consisted of actors representing the people affected by the simulated disaster: utility workers, industry representatives, police, firemen, a radio station, emergency services representatives, *etc.* The actors were responsible for simulating the disaster with all its parameters and information sources. Events were planned and logged in an electronic exercise planning system. Events were simulated using conventional media, such as press conferences, radio bulletins, press statements, phone calls and e-mail.

The second group consisted of teams of decision-makers who had to cope with the simulated disaster. The team that we observed worked in two designated rooms, interacting with each other and other teams through several artefacts that mediated the flow of information. We observed the decision-makers and not the actors.

The collocated team of decision-makers had access to the following artefacts:

- whiteboards with pens, in both rooms
- computers – portable and desktop computers with internet connections, e-mail and the PGCS
- phones – landlines and mobile phones in both rooms
- maps – electronic and hand-drawn maps
- newspaper – an electronic newspaper with relevant exercise information of the days before the exercise
- radio – a radio station broadcasting live coverage of the simulated disaster. The broadcast included news bulletins, press conferences, interviews, music, *etc.*

6 Results

This section presents a brief description of how the team worked in its collocated setting using a framework adapted from a taxonomy of team performance (Fleishman and Zaccaro, 1992). We relate the team functions to the artefacts used during the exercise and discuss the characteristics of the artefacts in relation to work in collocated and distributed settings.

6.1 Collocated work

The team usually works in a distributed manner. They typically stay at their home offices and coordinate their work through telephone conferences during which they discuss and present updates on the developing situation. The team has a meeting location for highly critical emergencies, but so far the members have never used it. They are openly sceptical of the possibility that they could ever work together in a collocated manner (*e.g.*, in their meeting location), since their companies do not want them to be out of their offices for long periods of time. Thus, the exercise setting did not match the actual work practice.

During the exercise, the team was collocated and had two rooms at its disposal. They made one room their meeting room. The other room was their hub for telephone calls and internet communication. Team members were initially assigned to one room or the other. The team met every hour in the meeting room to listen to the radio broadcast, to share information and to plan. These meetings were coordinated by the chairperson and the team members reported one at a time. Between meetings, most team members used telephones to contact other teams of decision-makers to exchange information. Occasionally they walked around, formed subgroups and discussed issues together. As the exercise developed, team members started to move back and forth between subgroups and rooms. Several times, the hallway between the rooms became a meeting point and workplace. When the situation was relatively calm, conversations were often more social than professional. The evolving informality and the fluid nature of the subgroups and room assignments are characteristic benefits of collocated work.

6.2 Team functions and artefact use

As the first step in our analysis, we identified seven categories of team performance functions. The categories draw upon a taxonomy of team performance developed by Fleishman and Zaccaro (1992). The seven categories (presented in the leftmost column of Table 1) cover the full range of managerial activities one expects to find in team work. The motivation for using Fleishman and Zaccaro's (1992) model is that it has been applied successfully to the analysis of team functions and that it covers the activities and functions of the studied team well. Other models of team performance have been developed, as discussed in a review by Militello *et al.* (1999). Fleishman and Zaccaro's (1992) model provides a well developed and rather comprehensive set of dimensions of team level performance, which makes it useful and applicable for the purposes of this study.

Table 1 Team functions (adapted from Fleishman and Zaccaro, 1992) and relative frequency of artefact use. The last column illustrates the frequency of face-to-face communication to mediate information to perform the functions

Team function	*Whiteboard*	*E-mail*	*PGCS*	*Phone*	*Maps*	*Newspaper*	*Radio*	*Face-to-face*
1. Orientation functions								
a. Internal info exchange, team resources and constraints	Continuous							Continuous
b. Internal info exchange, team task and goals	Continuous							Continuous
c. Internal info exchange, emergency management resources and constraints	Continuous		Continuous		Occasional	Seldom		Continuous
d. External info exchange, emergency management resources and constraints		Occasional	Continuous	Continuous			Frequent	
e. Team task priority assignment	Continuous							Frequent
2. Resource distribution functions								
a. Matching team resources to requirements	Occasional		Continuous					Frequent
b. Load balancing	Continuous							Frequent
3. Timing, pacing functions								
a. General activity pacing	Frequent							Frequent
b. Individual activity pacing	Seldom							Frequent
4. Response coordination functions								
a. Response sequencing	Occasional							Frequent
b. Timing and positioning responses	Frequent							Frequent
5. Motivational functions								
a. Development of performance norms								Seldom
b. Generating acceptance of performance norms								
c. Establishing team-level performance rewards								
d. Reinforcement of task orientation	Frequent							Continuous
e. Balancing team and individual focus								
f. Internal resolution of conflicting information	Frequent		Continuous					Occasional
g. External resolution of conflicting information		Occasional	Continuous	Continuous			Seldom	
6. Systems monitoring functions								
a. General activity monitoring	Continuous							Frequent
b. Individual activity monitoring								Frequent
c. Adjustment based on monitoring	Frequent							Frequent
7. Procedure maintenance								
a. Monitoring general procedure activities								Occasional
b. Monitoring individual procedure activities								
c. Adjustment of non-standard activities								

Since the team worked in a collocated manner, they had shared access to the artefacts in the rooms to perform these functions. We estimated the relative frequency with which they performed each function using the various artefacts. Face-to-face communication was included in this analysis to highlight its salient role in collocated work. In Table 1, we use the following terms to describe these frequencies: *Continuous* means many times per hour during the entire exercise; *Frequent* means about once an hour; *Occasional* means several times during the exercise; *Seldom* means once or twice during the exercise.

Team members performed each of the seven categories of functions to some extent:

1 *Orientation functions* include all information exchange about resources, constraints and task goals and priorities. Unlike Fleishman and Zaccaro, we distinguish between internal and external resources and constraints, and between internal and external information exchange. By 'internal' we mean within the team, and by 'external' between the team and other participants in the exercise. While face-to-face communication dominated internal information exchange, the major artefact used to mediate this function was the whiteboard. In contrast, ICT mediated external information exchange.

2 *Resource distribution functions* concern the matching of team resources to requirements and workload balancing. These functions were continuously performed using the PGCS, the whiteboard and face-to-face communication.

3 *Team activity timing and pacing* was performed frequently, mainly through face-to-face meetings and the whiteboard. It should be noted that team members' activity was not visibly paced with the help of ICT.

4 *Response coordination functions* were most frequently performed through face-to-face meetings and the whiteboard, and not through ICT.

5 We observed the team engage in several of the *motivational functions* enumerated by Fleishman and Zaccaro. The reinforcement of task orientation was performed continuously through face-to-face communication and frequently with the aid of the whiteboard. ICT was instrumental in the resolution of conflicts with people outside the room. Internal conflicts were resolved using the whiteboard, the PGCS and face-to-face communication. Typical conflicts concerned mismatches between what the team knew and what other participants stated to be true. The development of performance norms was an infrequent activity.

6 *Monitoring and adjustment* of general and individual activities was done frequently using the whiteboard and face-to-face-communication. ICT was not used to monitor or adjust team activities.

7 There was very little observable *maintenance of procedures*. The only procedures that applied and were followed were how the team's joint responses were supposed to be performed. This function was occasionally discussed face-to-face in team meetings.

As shown in Table 1, the whiteboard and face-to-face communication were the most frequently used media for all categories of team functions. The whiteboard and face-to-face communication can be used only when the team is collocated. When comparing the communication media (including ICT artefacts), it becomes clear that the remaining artefacts lack certain important characteristics for the utility teams'

communication. Being collocated enables a shared local context, informal interactions between team members, gestures, coreference to objects and implicit cues about peripheral activity (Olson and Olson, 2000). These characteristics clearly seem to be beneficial given the frequent use of face-to-face and whiteboard communication during the exercise.

7 Implications

We began the paper with rationale for three claims:

1 The team should benefit from being able to communicate face-to-face and develop trust and common ground.

2 They can develop patterns of interaction that may transfer to distributed work and, hence, enhance operations.

3 However, the communication, task allocation and artefact (ICT) use during training may not readily transfer to their actual distributed work setting.

The observational study sheds light on the first and the last of the three claims. We have no data to document any transfer of benefit to the team's actual distributed work setting. The frequent face-to-face communication and whiteboard use strongly suggest that the team benefited from the collocated setting. The continuous shared referencing to the whiteboard and the PGCS reveals that the team developed a consistent pattern of interaction that guided virtually all emergency management activity. Moreover, the emergence of fluid subgroups and social (non-professional) conversations is consistent with the development of mutual trust and common ground. None of these social aspects were mediated by ICT. Our observations suggest that ICT can be useful for the practice of emergency management. However, it is difficult to imagine that ICT would be used to mediate social interaction during distributed emergency management.

Table 2 summarises the possibilities for ICT use in collocated and distributed work settings and illustrates the potential mismatch between the training setting and distributed work. As noted above, whiteboards and face-to-face communication can only be used when the team shares the same physical location. The observation that these are the most frequent modes of communication by the emergency management team suggests that there may be a mismatch between what the team learned during training and what they need on the job. In addition, because tasks were assigned given the collocated work conditions, task allocation in training may not correspond to or be appropriate for actual distributed work. Concerning this mismatch, either the training mode should change or, during an actual emergency, the team should work together in its meeting location.

In the future, distributed work settings may feature artefacts such as videoconferencing and virtual (electronic) whiteboard facilities. These ICT artefacts may partially address the need to co-reference objects and other subjects of conversation. However, informal interactions such as socialising may be resistant to mediation by ICT. Since such interactions are instrumental for the development of trust and common ground, distributed teams may be at a serious disadvantage that technology may not be able to overcome.

Table 2 The impact of work setting and function on ICT use

Function	*Collocated setting*	*Distributed setting*
ICT use to communicate within the team	PGCS	PGCS
	Maps	Maps
	Newspaper	Newspaper
		E-mail
		Phone
ICT use to communicate with others	PGCS	PGCS
	Maps	Maps
	Newspaper	Newspaper
	E-mail	E-mail
	Phone	Phone
Other communication media	Whiteboard	
	Face-to-face	

Training for distributed emergency management poses a dilemma. On the one hand, collocated exercises have clear social benefits. On the other hand, the lessons and skills learned during collocated work may not transfer to the actual distributed work setting. We believe that the team would benefit from participating in more collocated exercises in the future. However, it is important that the ICT employed in collocated training closely resemble that used in their distributed work. If the training environment includes irrelevant ICT, some of the knowledge and work routines acquired during the training may be neither useful nor transferable to actual emergency management. The whiteboard and face-to-face meetings offer ways of communicating not afforded by ICT in a distributed setting. This constitutes a significant difference between the training and real settings. Because the communication media that were used most extensively during training are unavailable in the team's distributed work environment, the extent to which the training can generalise is unclear.

It is hard to see how this type of exercise can be performed other than in a collocated manner. To organise a realistic, large-scale, multi-organisation emergency management exercise is extremely difficult; to try to stage the scenario in a distributed fashion is even more difficult. Even though the observed exercise differed markedly from how the team usually works, the benefits of the exercise for future emergency management are clear. The experience of working in the same physical location most likely helps team members get to know each other better, which in turn can lead to the development of trust, common ground and more efficient communication. Exercises like the one we observed may also enable emergency management workers to discover new ways of working that can be incorporated in their organisation of work.

The utilities' exercise provides an example of how companies that provide services critical to society can set aside their commercial rivalry in order to improve their emergency management performance. The mismatch between the use of ICT during the exercise and actual work illustrates the challenges faced by ICT designers and developers of emergency management training exercises.

Acknowledgements

The authors gratefully acknowledge the editors, two anonymous reviewers, Erik Hollnagel and Kjell Ohlsson for useful comments on earlier drafts, Kjell Mo for giving us the opportunity to study the exercise, and ConDesign in Linköping for technical assistance.

References

Alexander, D. (2000) 'Scenario methodology for teaching principles of emergency management', *Disaster Prevention and Management*, Vol. 9, pp.89–97.

Alexander, D. (2003) 'Towards the development of standards in emergency management training and education', *Disaster Prevention and Management*, Vol. 12, pp.113–123.

Alexandersson, H. (2005) 'Den stora januaristormen 2005 [The big January 2005 storm]', *Väder och Vatten*, Vol. 1, p.11.

Baldwin, R. (1994) 'Training for the management of major emergencies', *Disaster Prevention and Management*, Vol. 3, pp.16–23.

Barron, J. (2003) 'The blackout of 2003: the overview; power surge blacks out northeast, hitting cities in 8 states and Canada; Midday shutdowns disrupt millions', *The New York Times*, New York, NY, 15 August.

Brown, J. (2004) 'Providing emergency care during a power outage: August 2003', *Disaster Management and Response*, Vol. 2, pp.20–23.

Clark, H.H. (1996) *Using Language*, Cambridge, MA: Cambridge University Press.

Cramton, C.D. (2002) 'Attribution in distributed work groups', in P.J. Hinds and S. Kiesler (Eds.) *Distributed Work*, Cambridge, MA: MIT Press.

Fleishman, E.A. and Zaccaro, S.J. (1992) 'Toward a taxonomy of team performance functions', in R.W. Swezey and E. Salas (Eds.) *Teams: Their Training and Performance*, Norwood, NJ: Ablex Publishing Corporation.

Hinds, P.J. and Kiesler, S. (Eds.) (2002) *Distributed Work*, Cambridge, MA: MIT Press.

Hollnagel, E. and Woods, D.D. (2005) *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*, Boca Raton, FL: CRC Press.

Hutchins, E. (1995) *Cognition in the Wild*, Cambridge, MA: MIT Press.

Jarvenpaa, S.L. and Leidner, D.E. (1999) 'Communication and trust in global virtual teams', *Organization Science*, Vol. 10, pp.791–815.

McAllister, D. (1995) 'Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations', *Academy of Management Journal*, Vol. 38, pp.24–59.

McEntire, D.A. (2002) 'Coordinating multi-organisational responses to disaster: lessons from the 28 March 2000, Fort Worth tornado', *Disaster Prevention and Management*, Vol. 11, pp.369–379.

Meyerson, D., Weick, K.E. and Kramer, R.M. (1996) 'Swift trust and temporary groups', in R.M. Kramer and T.R. Tyler (Eds.) *Trust in Organizations: Frontiers of Theory and Research*, Thousand Oaks, CA: Sage.

Militello, L.G., Kyne, M.M., Klein, G., Getchell, K. and Thordsen, M. (1999) 'A synthesized model of team performance', *International Journal of Cognitive Ergonomics*, Vol. 3, pp.131–158.

Norman, D. (1993) *Things that Make Us Smart: Defending Human Attributes in the Age of the Machine*, Perseus Books.

Olson, G.M. and Olson, J.S. (2000) 'Distance matters', *Human-Computer Interaction*, Vol. 15, pp.139–178.

Olson, J.S., Teasley, S., Covi, L. and Olson, G.M. (2002) 'The (currently) unique advantages of collocated work', in P.J. Hinds and S. Kiesler (Eds.) *Distributed Work*, Cambridge, MA: MIT Press.

Orton, J.D. and Weick, K.E. (1990) 'Loosely coupled systems: a reconceptualization', *Academy of Management Review*, Vol. 15, pp.203–223.

Perrow, C. (1984) *Normal Accidents: Living with High-Risk Technologies*, New York, NY: Basic Books, Inc., Publishers.

Perry, W.R. and Lindell, M.K. (2003) 'Preparedness for emergency response: guidelines for the emergency planning process', *Disasters*, Vol. 27, pp.336–350.

Regester, M. and Larkin, J. (2005) *Risk Issues and Crisis Management: A Casebook of Best Practice*, 3rd ed., London, UK: Kogan Page, Limited.

Säljö, R. (2000) *Lärande i Praktiken: Ett Sociokulturellt Perspektiv [Learning in Practice: A Socio-cultural Perspective]*, Stockholm: Prisma.

Schunn, C., Crowley and Okada, T. (2002) 'What makes collaborations across a distance succeed? The case of the cognitive science community', in P.J. Hinds and S. Kiesler (Eds.) *Distributed Work*, Cambridge, MA: MIT Press.

Suparamaniam, N. (2003) 'Renegotiation of authority in the face of surprise', *Unpublished Doctoral Dissertation*, Linköpings universitet, Linköping, Sweden.

Turoff, M. (2002) 'Past and future emergency response information systems', *Communications of the ACM*, Vol. 45, pp.29–32.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Games and Economic Behavior 52 (2005) 257–282
GAMES and Economic Behavior
www.elsevier.com/locate/geb

A brain imaging study of the choice procedure

Aldo Rustichini [a], John Dickhaut [b], Paolo Ghirardato [c,*], Kip Smith [d], Josè V. Pardo [e]

[a] *University of Minnesota, Department of Economics, 1268 Heller Hall, 271 19th A. So., Minneapolis, MN 55455, USA*
[b] *University of Minnesota, Carlson School of Management, Department of Accounting, Minneapolis, MN, USA*
[c] *Dipartimento di Statistica e Matematica Applicata and ICER, Università di Torino, Piazza Arbarello 8, 10122 Torino, Italy*
[d] *Kansas State University, Department of Psychology, Bluemont Hall 492 1100 Mid-Campus Drive, Manhattan, KS 66506-5302, USA*
[e] *University of Minnesota, Veterans Affairs Medical Center, MN, USA*

Received 8 June 2004

Available online 10 December 2004

Abstract

We study the behavior of subjects facing choices between certain, risky, and ambiguous lotteries. Subjects' choices are consistent with the economic theories modeling ambiguity aversion. Our results support the conjecture that subjects face choice tasks as an estimation of the value of the lotteries, and that the difficulty of the choice is an important explanatory variable (in addition to risk and ambiguity aversion).

The brain imaging data suggest that such estimation is of an approximate nature when the choices involve ambiguous and risky lotteries, as the regions in the brain that are activated are typically located in parietal lobes. Thus such choices require mental faculties that are shared by all mammals, and in particular are independent of language. In contrast, choices involving partial ambiguous lotteries additionally produce an activation of the frontal region, which indicates a different, more sophisticated cognitive process.
© 2004 Elsevier Inc. All rights reserved.

* Corresponding author.
E-mail addresses: arust@econ.umn.edu (A. Rustichini), jdickhaut@csom.umn.edu (J. Dickhaut), ghirardato@econ.unito.it (P. Ghirardato), kip@ksu.edu (K. Smith), jvpardo@james.psych.umn.edu (J.V. Pardo).

0899-8256/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.geb.2004.08.005

JEL classification: D81

Keywords: Brain imaging; Decision theory; Ambiguity; Procedural choice

1. Introduction

"*The mental operations by which ordinary practical decisions are made are very obscure, and it is a matter of surprise that neither logicians nor psychologists have shown much interest in them* (Frank Knight, 1921, Chapter 7)".

1.1. Risk and uncertainty

In *Risk, Uncertainty and Profit*, Frank Knight (1921) introduced a distinction between risk and uncertainty. He called a choice environment *risky* if the outcome is random, but the person making the choice can reasonably attach a numerical probability to each event. He called it *uncertain* if the subject cannot define uniquely and precisely such numerical probability. For example, if the probability of the events are described "objectively," [1] as in tossing a coin or rolling a die, then it is natural to expect that the subject will attach probabilities to events equal to the objective ones. On the other hand, in a situation of uncertainty the subject may have no reasonable estimate of the frequency of the events, say, because the event or choice he is asked to consider is a of a unique, once-and-for-all nature.

The basis of the difference derives in Knight's view from the way in which estimates of probabilities are derived. In the case of risk, the probability distribution over possible outcomes is known. This may be due to an a priori calculation (as in the case of the roll of a die) or frequency estimation. On the other hand, this knowledge is not available in the case of uncertainty, either because no obvious list of equally likely and exhaustive basic alternatives is available or "because the situation is in high degree unique" (Knight, 1921, Chapter 8).

This point of view may seem outdated in view of the rise, in the years to follow Knight's book, of the subjectivist approach to probability of Ramsey, de Finetti and Savage.[2] In the classical theory of choice under uncertainty called "Subjective Expected Utility" (SEU for short)—the most complete treatment of which is the axiomatic formulation by L.J. Savage (1954)—such distinction vanishes. In SEU theory, it is assumed that the subject is able to provide a subjective probabilistic estimate of the relative probability of each event. Once he has done that, the evaluation of a lottery (or, more generally, a state-contingent payoff function) is the same under risk or uncertainty: It is the expected value, computed with respect to this probability. Therefore, in terms of the Knightian distinction, according to SEU theory all uncertainty can be reduced to risk.

[1] Knight (1921, Chapter 8) states that the terms objective and subjective are equivalent to those of risk and uncertainty, but this identification may be confusing today.

[2] The first draft of the book was Knight's doctoral dissertation at Cornell, 1915–1916.

However, in their actual behavior human subjects may fall short of the expectations of theorists. Indeed, among scholars of economics and decision making there has been a resurgence in the interest on Knight's distinction motivated by a very interesting critique of SEU theory formulated by Daniel Ellsberg (1961). Moreover, we will argue in this paper that the distinction may acquire more interest in light on the surging attention on the mental process that the subject follows to formulate his estimate of the likelihood of events.

1.2. Ellsberg's "paradox"

Ellsberg (1961) begins with the Knightian distinction, using the now more customary term *ambiguity* instead of Knight's "uncertainty." Rather than trying to base the distinction on the way in which the probabilities are estimated, he accepts the purely subjective view of Ramsey, de Finetti and Savage: "The degree of a belief is the extent to which we are prepared to act upon it" (Ramsey, 1926). That is, whether subjects are able to attach numerical probability or not to two events can only be measured by their choices over acts based on these events. If subjects are willing to give us an answer to every choice we propose them, and provided that their answers are consistent, this measurement is unambiguously defined.

His classical thought experiment is the first attempt to test whether there is a significant difference between ambiguity and risk. In his experiment, subjects face the choice among lotteries, where the outcome is described by draws of balls from an urn. For an urn with Red, Black and Yellow balls, I can define a lottery as three numbers, assigning, say, a monetary amount to each of the three outcomes. The proportion of the balls, however, is not completely specified: For example, in his classical design, an urn has 90 balls, of which 30 are Red and 60 are Black and Yellow, with the relative number of Black and Yellow balls unspecified. So the subject is not provided with an objective probability: Does he always provide his own, well-specified, subjective probability over the different outcomes, as SEU theory requires?

Ellsberg's experiment:[3] The experimenter tells the subject that an urn has the composition of Red (R), Black (B) and Yellow (Y) balls as described above. Then, the subject is asked to choose between the lotteries a and b whose payoffs (in dollars) are determined according to the following table:

	R	B	Y
a	100	0	0
b	0	100	0

After the subject has made this choice, the experimenter asks him to choose between the lotteries c and d defined by the following:

[3] Though Ellsberg by his own admission tried his experiment under "absolutely non-experimental conditions," this pattern of choices has been observed in a multitude of properly conducted experiments. See, e.g., Luce (2000) for references to this vast literature.

	R	B	Y
c	100	0	100
d	0	100	100

Most subjects choose a in the first pair of lotteries, and d in the second. This is a violation of SEU theory. The intuitive reason is clear: If you are a SEU decision maker, and choose a in the first choice you must think that the outcome Yellow ball is more likely. But then this fact should make you choose c, not d in the next choice.

Thus, the typical subject of Ellsberg's experiment does not form a subjective probability for each event, even upon reflection, and even after interrogation or prodding by unsympathetic critics.[4] In fact, if the assumption that the subject estimates the different alternatives on the basis of some form of expectation is maintained, his choices show that such unique subjective probability cannot exist. When this occurs, we say that the subject's choices display *ambiguity aversion* (*or love*). Clearly, it is the absence of a well-defined objective probability in the experimental design that provides the conditions for ambiguity aversion/love to manifest itself.

The inability of subjects to form an estimate of the probability of different events was, until recently, a controversial issue mainly for scholars in economics. It may acquire larger interest now that neuro scientists have entered into this specific arena. A basic assumption of their program is that the two fundamental operations defining expected utility (the estimate of a probability and the estimate of value) have a neural basis. These are physical computations, performed by neurons. If subjects reveal with their choices that a unique probability does not exist, how can this operation have a neural basis? What is ambiguity aversion telling us about the psychological and neurological processes underlying decisions? Note that subjects in Ellsberg's experiments typically *do* choose one of the lotteries. So some decision process must have taken place. If the outcome is not consistent with the evaluation of expected utility with respect to a probability distribution, what was the process? For example, when Ellsberg introduces ambiguity as an explanation of his experimental observations, he explicitly notes it as a third component, in addition to probability and value, in the evaluation of a possible action.[5]

In light of the experimental robustness of ambiguity aversion, economists have developed extensions of SEU theory which incorporate this third component in a subject's decision rule. One of the most popular extensions is the so-called "Maxmin Expected Utility with multiple priors" (MEU for short) model of Gilboa and Schmeidler (1989). According to this model, the subject's beliefs are given by a *set* of probabilities (equivalently, the subject's beliefs on each event are given by an interval, rather than point, estimate).

[4] These are called "deliberate violators" by Ellsberg (1963). Among them, Ellsberg reports, L.J. Savage.

[5] He says (Ellsberg, 1961, p. 657):

"Responses from confessed violators indicate that the difference is not to be found in terms of the two factors commonly used to determine a choice situation, the relative desirability of the possible payoffs and the relative likelihood of the events affecting them, but in a third dimension of the problem of choice: the nature of one's information concerning the relative likelihood of the events."

The subject chooses the action whose worst-case expected utility evaluation (the minimum expected utility ranging over the set of possible probabilities) is highest.[6]

1.3. The psychological nature of ambiguity aversion

Ellsberg's paradox suggests that the Knightian distinction is substantial—and moreover that not all uncertainties can be reduced to risk—by proving that there is a difference in the behavior of subjects in choices under risk and choices under uncertainty. While the existence of the difference seems demonstrated, some still doubt its substance and importance. For instance, they argue that subjects might be confused. Deliberate violators might give up their violations facing a tighter experimental design (for instance, with an urn placed in front of their eyes during the entire experiment). Or, in the experimental situation in which ambiguity aversion appears, subjects might be simply exhibiting mistrust toward the experimenter.

The question we address in this paper is to determine experimentally whether there is a fundamental psychological difference between the two. We do this by analyzing the decision process; i.e., the sequence of different activities that are performed in taking a decision.

1.4. Decisions and emotions

Evidence from neuroscience suggests that choices under risk and ambiguity might be fundamentally different from the psychological point of view. In a series of classical papers and books, Antonio Damasio and his group have suggested that emotional components enter ambiguous choices in a crucial way. The evidence for this statement, which is reviewed more in detail later, is both clinical and experimental.

The clinical evidence is provided by a set of human subjects with lesions in the prefrontal cortex. These subjects are well known to have difficulties both in expressing and forming emotions, as well as in taking decisions. However they are not in any significant way impaired in their intellectual, cognitive and memory abilities.

Human subjects choosing a deck out of a set of card decks provide the experimental evidence. A payoff, real or hypothetical, is associated with each card, according to a probability that is not completely specified to the subjects. They are asked to choose one of the card decks; once they choose, they draw a card from the deck, examine the card, and discover the payoff of that card. Then, they proceed to the next choice. Just as in the case of the Ellsberg's urn, the experimenter does not provide the subject with a complete description of the stochastic process he is facing. The truth is that some of the decks have higher gains, but also higher losses, while others have lower gains as well as lower losses. The expected value of the first is lower than the second. The observation is that normal subjects tend over time to switch to the safer decks, while the choices of patients with frontal lesions converge to the first. Test of the emotional reactions to choices are also revealing: in the moments preceding the choice of the risky decks normal subjects show an active Galvanic

[6] In Ellsberg's experiment, a subject whose set of probabilities is given by all Ps such that $P(R) = 1/3$ and $P(B) \geqslant 1/4$ would make the typical ambiguity averse choices.

Skin Response (a measure of emotional reactions), while patients do not. The conclusion is also supported by brain imaging studies of subjects while making choices in the card decks experiment. Decision making, as Antoine Bechara and Antonio Damasio conclude (Bechara and Damasio, 2003), is a process driven by emotions.

1.5. Choices and emotions

On the other hand, subjects in the card deck experiment are performing several tasks at the same time they are choosing between alternatives: but they are also learning about the environment that they are facing, and they are receiving, after every choice, a feedback on their wins or losses. The involvement of emotional factors might therefore occur for any of these three different reasons.

Our experimental design aims at separating the choice from the learning and the payoff feedback, and tests the processes that are active in the presence of ambiguity. Our conclusion is that choice is a process driven by cognitive factors, even within those subjects that in the experiment unambiguously display ambiguity aversion.

1.6. Our experiment

The long-term aim of the research reported here is to identify and test a theory of how subjects actually reach their decisions. We think it is advisable, in this first phase of the research, to focus on the analysis of simple decisions, based on the choice between pairs of economic stimuli. Consequently, we chose a decision problem in line with the original Ellsberg's thought experiment (Ellsberg, 1961). We did not expect the choice behavior of subjects to be different from that predicted by existing "as if" theories of choice under risk and ambiguity, like the MEU model mentioned above. In fact, the analysis of the choice data in section below shows that they were not.

A second element in our choice of design was the introduction of a *partially* ambiguous lottery. In a risky lottery the subject knows the objective probability of outcomes, in an ambiguous lottery he has no information on this probability. In a partially ambiguous he has *some* information. The partially ambiguous lottery is located, *from a choice-theoretic point of view*, in an intermediate position between risky and ambiguous lottery. We will argue, however, that from a procedural point of view it has a very specific nature, so that the behavior of a subject facing a partially ambiguous lottery is very different. This is indeed what the analysis of the data on response times and the imaging data suggest.

1.7. Content of the paper

In Section 2 we describe the experimental design. In Section 3 we present and discuss the behavioral data. More precisely, in Section 3.1 we examine the choices made by the subjects in different conditions, while in Section 3.2 we focus on the response time; namely the time used by the subject to reach each decision. In Section 4 we present and interpret the brain-imaging data in the light of information available in the neuroscience literature on the significance of the different patterns and centers of neural activation. Finally, in Section 5 we draw our conclusions and outline the future research agenda.

2. Experimental design

Subjects were instructed to make a sequence of choices between pairs of lotteries. The pairs were presented in groups of similar choices, and no feedback on the outcome was provided during the test. Outcomes and payments were determined at the end.

2.1. Lotteries

In the entire experiment, four different types of lotteries were used: certain (C), risky (R), partially ambiguous (PA) and ambiguous (A) lotteries. Subjects were informed that the payoff to such lotteries would eventually be determined by the draw of a ball, which could be either blue or red, from an urn containing 180 balls. The number of balls of each color was to be consistent with the information given about the lottery, as specified below. Overall subjects had to make 96 choices. The actual payments were decided at the end of the experiment. First, 4 out of the 96 choices were randomly selected according to a uniform distribution.[7] We then checked which of the two lotteries had been selected in these choices, filled a real urn with balls consistently with the subject's information and asked the subject to pick one of the balls, while keeping the urn above his/her head. The subject was then paid the total of the payments for the four choices.

The pair of lotteries in each choice was presented on a screen, indicating the number of balls for each color and the amount in dollars that (a draw of a ball of) each color would pay. The only exception was the certain lottery, for which the screen simply indicated a fixed amount in dollars. Subjects knew that the urn always contained a total of 180 balls. In the R lottery they were told that the urn would contain 90 balls of each color. For the A lottery no information on the number of balls of either color was given (only that the balls could only be red or blue). Finally, for the PA lottery they were told that the urn would contain at least 10 balls of each color, leaving the composition of the remaining 160 unspecified (but again, that they could only be blue or red).[8]

2.2. Choices

In each choice, it is useful to classify one of the two lotteries as the *main* lottery and the other as the *reference* lottery. The main lottery was either risky, partially ambiguous or ambiguous. Thus, subjects were faced with increasing levels of ambiguity: From no ambiguity in the risky lottery to full ambiguity in the ambiguous one. The main lottery was to be compared to the reference lottery, which could be either risky or certain. We used all possible combinations of main and reference lotteries to obtain six types of choices, the *conditions* in our experiment. Each condition is denoted by the type of its lotteries: For example, the condition PAC faces the subject with a choice between a partially ambiguous

[7] The small number aims at making each subject's choices close to his true preference over the lotteries involved. With a large number, a subject might use a strategy over the entire portfolio of choices that would make the optimal lottery in each choice different from the one he would select if facing that specific choice in isolation.

[8] For the sake of determining a subject's actual payoff, the actual urn compositions in the A and PA case were chosen randomly.

lottery (PA, the main lottery), and a certain lottery (C, the reference lottery). The condition AR faces him/her with the choice between an ambiguous and a risky lottery, and so on. Overall we had three conditions where the reference lottery was of the R type (the R-conditions RR, PAR, AR), and three where the reference lottery was of the C type (the C-conditions RC, PAC, AC).[9]

2.3. The lotteries

A detailed description of the different lotteries is provided in Appendix B. Here we point out some specific features of the set of choices we used, because understanding them is essential in the interpretation of the results.

In the C condition subjects are comparing a certain amount (ranging from a minimum of 10 dollars to a maximum of 50) with either a risky, partially ambiguous, or ambiguous lottery. In the R condition the reference lottery is a risky, rather than certain, lottery: this choice may appear more difficult, but not necessarily in the specific setup we adopted. In fact, the reference lottery in fact *dominates* the main lottery, in the following sense. In the RR choice, the main lottery is a mean-preserving (variance-increasing) spread of the reference lottery. For example, the main lottery has payoffs $(64, 0)$, while the reference lottery has payoffs $(60, 4)$, with an equal probability for each type of ball. In the AR and PAR conditions, this negative effect is compounded by the ambiguity associated with the main lottery. For example, the reference lottery has payoffs $(60, 4)$ with fifty–fifty probability, while the main lottery has payoffs $(64, 0)$ for red and blue balls respectively, with an unspecified proportion of red and blue balls.

The joint effect of risk and ambiguity should therefore make the choice of the main lottery look inferior to a subject who is averse to risk and ambiguity. In addition, this comparison should involve simple qualitative reasoning, rather than quantitative comparisons. On the other hand, the choice in the C conditions involves a quantitative comparison, since an estimate of the value of the main lottery is compared with the certain value of a C-type lottery. As we are going to see, this difference between the two R and C conditions is also suggested by our experimental observations.

2.4. Time sequence

Each subject experienced the six conditions (RC, PAC, AC, RR, PAR and AR) that we have just described, plus two with Eyes-Closed-Rest (ECR). The conditions and the set of choices in each condition were the same for each subject. The order in which the conditions were presented, and the order of choices within each condition, was determined randomly and independently for each subject.

[9] The names "main" and "reference lottery" are used here for expository purposes only. These names were never used in the experiment, and neither were the labels "certain," "risky," "partially ambiguous" and "ambiguous."

2.5. Imaging technique

The imaging study was conducted using the PET (Positron Emission Tomography) scanning technique. General information on the technique is given in Appendix C, together with a more detailed description of the technique used in the study.

2.6. Implementation

The original sample was composed of 12 young healthy right-handed individuals, chosen among those answering a public announcement posted on campus. One of the subjects had to be excluded from the sample after post experiment interviews determined a state of depression[10]; for a second the data on scanning were lost for technical reasons. Therefore, the data in this study refer to the sub-sample of 10 individuals.

Subjects came in separately, on different days. We first paid each subject 50 dollars in cash. This show-up award was never at risk during the experiment. We then read the instructions. The instructions were very detailed; we also asked the subjects to answer short quiz questions during the presentation to check their understanding. Detailed and careful instructions were intended to make the subject familiar with the four different types of lotteries and the six different conditions. We presented a set of examples, and asked the subject to choose among the lotteries in the example. This also served the purpose of familiarizing the subjects with the method of expressing their choice, a click on the left or right button of a mouse.

After the instructions, the subjects were moved and were positioned in a scanner. Choices were made while the brain activity of the subject was scanned. We had 15 choices for each R condition and 17 choices for each C condition, for a total of 96 choices per subject. A choice appeared on the screen, and subjects had six seconds to decide. The time interval between choices was fixed, and independent from the moment in which the choice was made. A pause of two seconds would follow the end of each choice, and then the next choice would be displayed on the screen (so the overall time interval between choices was eight seconds). The time interval between the different conditions varied between two to four minutes, since a new condition could begin only when the scanner was ready for the next analysis. For each subject, the entire experiment lasted approximately two hours.

3. Behavioral data

3.1. Choices

3.1.1. The C conditions

The observed choices in the C conditions tend to follow a rather regular cutoff policy. Each subject chooses the R, PA or A lottery rather than the C lottery when the certain amount is below a threshold (which varies with the subject), and switches to the C lottery when the threshold is passed.

[10] This is standard procedure: the data from depressed subjects are not considered reliable.

Table 1
Summary statistics for the cutoff in the C conditions

Variable	Obs.	Mean	Std. err.	95% conf. int.
AC cutoff	170	22.7	0.439	[21.83, 23.56]
PAC cutoff	165	21.7	0.3769	[20.95, 22.44]
RC cutoff	153	28.11	0.217	[27.68, 28.54]

Table 2
Choices in the C conditions

Subject	27	29	40	44	52	53	55	59	68	71	Average	Median
Cutoff in RC	25	25	25	33	31	30	28	28	28	*	28.11	28
Deviations from cutoff	0	0	0	2	3	3	3	0	1	*	1.33	
Cutoff in PAC	20	25	15	32	20	30	20	20	15	25	21.7	20
Deviations from cutoff	0	0	0	1	2	0	2	1	0	0	0.6	
Cutoff in AC	20	20	15	31	30	28	20	28	15	20	22.7	20
Deviations from cutoff	0	0	0	2	0	0	0	0	0	1	0.3	
PAC–AC	0	5	0	1	−10	−3	0	−8	0	5	−1	

* Denotes missing data.

Estimates of the cutoff point are summarized in Table 1.[11] The cutoff value is chosen for each subject so as to minimize the number of deviations, for that subject, of the observed choices from the cutoff policy.[12]

Table 2 shows that subjects were consistent in their choices, and that the instances of deviations from the policy implicitly described by the cutoff are small in number.[13]

The bottom row of Table 2 reports the differences in the value of the cutoff for PAC and AC conditions. The differences are zero or small: this indicates that the choices of the same subject are consistent across conditions. The "Mean" column of Table 1 and the last row of Table 2 show that the values of the cutoffs in the two conditions PAC and AC are similar.

3.1.2. The R conditions

Table 3 reports the number of times each subject chose the risky reference lottery in the R conditions. Subjects chose the riskier lottery (the one with the greater spread) most frequently (but still only 14.7 percent of the times) in RR, and less frequently in PAR and AR (approximately the same in the two conditions).

3.1.3. Summary of the analysis of choices

Overall, the observed choices of the subjects are those predicted by widely accepted theories of choice in risky and ambiguous environments, like the MEU model. Between

[11] Some of the data are missing because either the subject did not choose in the amount of time available, or because of an error in recording the answer, for subject number 71.

[12] More precisely, the cutoff has been determined according to the following rule. A c-policy is the policy of choosing the C lottery if its value is larger or equal to c. For each of the possible values of c, determine the number of deviations from the c-policy in the observed choices of the subject. Choose the c that minimizes the number of deviations. If the value of this c is among the values at which the subject expressed indifference, choose the middle if the number of such values is odd, and the next one in ascending order if the number is even.

[13] Subjects are indicated by the classification number in data archive of the Veterans Affairs Medical Center.

Table 3
Choices of the main lottery in the R conditions

Subject	27	29	40	44	52	53	55	59	68	71	total	%
RR	0	4	0	3	0	6	1	6	2	*	22	14.7
PAR	1	0	0	4	0	1	1	4	1	0	12	8
AR	1	2	0	2	2	3	1	3	0	0	14	9.3

* Denotes missing data.

two lotteries, where one is a mean-preserving spread of the other, the subjects chose consistently and almost exclusively the lottery with smaller variance. Subjects were also ambiguity averse. This is hard to detect in the R conditions where the choice is already almost entirely of the lottery with smaller variance. But in the C conditions, the mean cutoff is six to seven dollars higher when the main lottery is R than it is when the main lottery is A or PA (see Table 1).

3.2. Response times

The *response time* (RT) is the length, in 1000th of a second (msc), of the time interval between the moment in which the stimulus (the two lotteries) appears on the screen and the moment in which the subject clicks on the mouse making the choice. Tables 4 and 5 below present the first surprise. They show the average response time, taken over subjects and different choices in the same condition, together with some summary statistics.[14]

The response time is approximately half of a second (that is, 25 percent) longer in the R conditions than in the C conditions. Among the C conditions, the fastest decisions were made in the AC and PAC conditions. The slowest decisions were made in the corresponding R conditions, namely AR and PAR. This disparity in response time suggests that subjects approached the two conditions with different mental processes.

Table 4
Average response times (RT) in the R conditions

Variable	Obs.	Mean	Std. err.	[95% conf. interval]
RT in AR	147	2776.95	87.24	[2604.52, 2949.37]
RT in PAR	165	2741.74	94.58	[2554.85, 2928.64]
RT in RR	148	2723.27	92.00	[2541.45, 2905.10]

Table 5
Average response times (RT) in the C conditions

Variable	Obs.	Mean	Std. err.	[95% conf. interval]
RT in AC	170	1947.60	65.98	[1817.34, 2077.85]
RT in PAC	165	2196.72	76.32	[2046.01, 2347.42]
RT in RC	153	2534.43	84.38	[2367.70, 2701.16]

[14] The number of observations is different across conditions. This happens for two reasons. First, some of the observations were lost for technical reasons. Second, the number of choices in the R condition were 15, and 17 in the C condition.

3.2.1. Difficult decisions and learning

Several factors may affect the length of time a subject needs to make a choice. Some insight into the determinants of this time (and hence on the decision process itself) can be obtained by a simple regression. A detailed report of these results is presented in Tables A.1 and A.2.

3.2.2. Learning in C

Consider first the C conditions. We use two variables. The variable *discut* is defined as the absolute value of the difference between the value of the certain lottery and the cutoff point that we have estimated for the subject. When the distance is very small, the subject is probably almost indifferent between the two alternatives, so in terms of the utility to the subject the decision is less important. On the other hand, the conclusion that s/he is indifferent is the outcome of a real life decision process, rather than its starting point. To reach this outcome, the subject might need less time when the value of the certain lottery is farther from the cutoff point, since in this case even an approximate estimate of the value of the main lottery will suffice. This finding, that response time increases as the certain value gets closer to the cutoff, is in line with what a procedural model of choice would predict.

A second variable is the integer-valued *order*, describing the order in which the choice has been presented to the subject in the same condition. If some form of learning takes place, then the response time will fall as the subject is facing choices that are becoming familiar.

The coefficient for the distance from the cutoff point (*discut*) is significant in the three C conditions, and has a negative sign. This is what we would expect to see if the task of deciding involves a significant comparison of two quantities, in our case the value of the certain lottery and some estimate of the value of the main lottery. This is also in agreement with the findings in purely cognitive studies. A strong non-linearity, with the response time increasingly in steep way as the term of comparisons are closer is well documented in cognitive psychology and neuro psychology (see for instance Pinel et al., 2001).

There are some interesting differences among conditions. Both *discut* and *order* have significant coefficients in the regression for the PAC condition. The coefficient for the variable *discut* is -56 msc per dollar, (with a p-value < 0.0001), the coefficient for the *order* variable is -21 per unit (p-value < 0.039). On the other hand, there is no significant difference in the latter coefficient if one estimates separately the initial and later choices. This indicates a regular, progressive learning, rather than a two-stage process—with an initial stage where subjects decide a policy in the form of a cutoff and a second stage in which they simply implement the policy. The *discut* variable has a significant coefficient in the AC and RC conditions as well, but the coefficient in AC is significantly smaller than in the PAC condition. The *order* variable is less significant, or insignificant, in the RC and AC cases respectively.

3.2.3. Learning in R

Here we consider three variables. The first is *value*, the expected value of the reference lottery, which ranged in the experiment between 30 and 40. The second is *order*, with the same meaning as in the previous section. The third and last is *variance*, a dummy variable with values $-1, 0, 1$ indicating the low, medium and high variance in the reference lottery.

Only the variable *variance* is significant, at least in the PAR and in the RR condition. The lack of learning is in agreement with the idea that the conditions where R is the reference lottery are easier. However, it makes the length of the response time in these very conditions even more surprising.

3.2.4. What operations do the subjects do?

The average value of the response time and the way it changes over the course of the trial can give some information on the type of operations subjects are performing. It is useful to compare our data with those for subjects performing a "pure" cognitive task.

In Spelke and Tsivkin (2001), the authors conduct a careful study of the response time for addition of two-digit integer numbers.[15] They study both approximate and exact operations. In the exact addition treatment subjects had to decide between the right answer and a distractor where the tens place was increased or decreased by 1. In the approximate addition treatment the problem was the same, but the candidate answers were multiples of 10, with the most distant answer 30 units more distant than the value closest to the correct answer. The average response time in both cases is (before training) between 4 and 4.5 seconds, a quantity much larger than we observe.[16]

On the other hand, the coefficient for the variable *discut* is large when compared to estimates of the effect of the difficulty of the problem induced by the proximity of the quantities to be compared. Consider for instance the finding in Pinel et al., 2001. In that study subjects had to perform a numerical comparison task: Specifically, they had to decide whether a visually presented number was larger or smaller than a fixed reference number, 65. The *numerical distance effect*,[17] namely the effect of the distance from 65 of the number presented to subjects on their response time, was estimated. The average response time was 600 msc for far numbers, slightly larger for moderately distant numbers, and 700 msc for the close numbers.[18]

4. Imaging results and analysis

4.1. Technical premise

We present the basic concepts necessary to understand the brain images. A more detailed explanation of the PET technique and of the statistical analysis underlying the study is given in Appendix A.

[15] For example, in the exact addition treatment the subjects had to add a first addend, which was ranging from 22 to 86, to a second addend ranging from 18 to 86 with the sum ranging from 40 to 172.

[16] No specific details are given in the study, but it seems that subjects had no time constraint.

[17] This effect is defined and discussed in detail in Dehaene et al. (1998). A second effect, the *number size effect*, was also documented in (Dehaene et al., 1998): For equal numerical distance, the discrimination of two numbers worsens as their numerical size increases.

[18] Numbers close to 65 were in the intervals 60–64 and 66–69; numbers moderately distant 50–59 and 70–79; numbers far 30–49 and 80–99. These times are much shorter than we observed: but the task of these subjects was a simple comparison of two numbers.

A point in the brain is defined by a triple of coordinates (x, y, z), with x the coordinate in the right to left direction, y the coordinate in the front to back direction, and z in the top to bottom direction. A positive x value denotes a position on the right, a positive y in the anterior part, and a positive z a position in the top part of the brain. The origin of this system of coordinates is roughly in the middle of the brain. Together, the triple (x, y, z) defines a point in a standardized three-dimensional model of the brain. The very small volume of brain around each such point is called a *voxel*.

Our observations are N vectors of rCBF (regional Cerebral Blood flow, Appendix A), one for each volume described by the three coordinates (x, y, z) in the brain of each of the N subjects. As different subjects have brains of different shape and size one of the first steps in data reduction to map the observations for the different subjects into a single standardized brain.

A statistical test is then used to estimate the probability that the different levels of rCBF in two conditions (for instance, in the PAC and the AC condition) at a specific point labeled by a triple (x, y, z) is different from zero. It is possible that two different conditions have a rCBF significantly different from the ECR condition, but also that the levels are so similar that the difference is not significant. The Z score is the statistic we use to report the probability that the difference is different from zero. The test is based on the assumption of normality and independence of the error, even in voxels that are very close.

There is a Z score for each voxel (and for each pair of conditions). The data can be more easily interpreted if a map of the difference score is presented in a picture.[19]

The images in the figures present the Z score for each voxel, associating different colors to different scores. First, only the voxels where the value of the Z score is above 2 are shown in color. A green color denotes a value between 2 and 3, yellow between 3 and 4, red between 4 and 5. All regions with value above 5 are white in color. In the images, the top part of each section corresponds to the front (rostral) part of the brain, the left part to the *right* part of the brain.

The values of the three coordinates are given here in millimeters (mm). The images show horizontal (also called transversal) sections of the standard brain, with the Z scores overlaid in color. The sections begin with the top and descend to the bottom. The numbers report the value of the z section, in millimeters. The standard model of the brain is that reported in the Talairach and Tournoux (1988) atlas.

4.2. The evidence from brain images

4.2.1. Overview

The activation is mostly in cortical areas, particularly frontal and parietal. There is no significant activation of areas (like the medial orbito frontal—or in general orbito frontal—and the limbic system, in particular the amygdala) that have been associated with the effect of emotions on decision making. The significance of this finding is discussed in detail in Section 5.3.1. The images support the idea that the procedure selecting the choice is mostly

[19] Colors are essential for the interpretation of the images, so a color printer is necessary. A copy of the images can be downloaded at http://www.econ.umn.edu/~arust/neuroecon.html.

of a cognitive nature, possibly involving some approximate computation (this hypothesis is discussed in detail in Section 5.2.2).

The R and C conditions are qualitatively different: the R conditions have modest activation compared to the C conditions. This finding supports the conjecture that the process involved in the choice in R conditions is simpler than the one in the C condition. These issues are discussed in detail in Section 4.2.2.

Among the C conditions, AC and RC differ from PAC. The first two have activations concentrated in parietal areas. The PAC condition has activations of the parietal and frontal areas. So the PAC condition plays a special role. In fact, the subtraction PAC–RC seems a weaker version of PAC–AC. This is particularly surprising in view of two facts. First, considered as a decision problem the difference between the AC condition and the PAC condition seems very small. The decision maker is told that the number of balls of each type can be anywhere in the interval $[0, 180]$, while in the second it can be anywhere in the interval $[10, 170]$, an apparently minor difference. Second, two sets of behavioral data suggest a similarity between PAC and AC as compared to RC. The cutoff point is in all subjects very close in the first two conditions, and rather different in the last. Also, the response times in the PAC and AC are similar, and different from the RC condition.

4.2.2. C conditions versus R conditions

As we observed above, the most active contrasts are in the C conditions. This is particularly true if one considers the difference between the various treatments and the ECR condition.[20] Among the C conditions, the most active is PAC. Similarly, among the R conditions the most active is PAR.

A large active region common to many of the differences between the C condition and the ECR is in the occipital lobe, lingual gyrus, with a peak around $(1, -75, 3)$. This region is for example active in PAC–ECR and RC–ECR. Interestingly, it is considerably less active in AC–ECR. This is the primary visual cortex (V1). The activity is due to increased visual attention. The higher activity in the C conditions is indirect evidence that this task induces a relatively greater amount of visual scanning of the main lottery for the purpose of defining the cutoff that is subsequently compared to the constant reference lottery.

4.2.3. The PAC condition

The two differences PAC–AC and PAC–RC have similar patterns. The main areas of activation in the two differences PAC–AC are:

(1) a region in the right frontal lobe, middle frontal gyrus, with peak at $(42, 50, -2)$, with a Z score 4.59;
(2) a region in the parietal lobe: in the subgyrus, with two peaks: one at $(25, -55, 42)$, with a Z score 4.42, and the other at $(34, -55, 33)$, with a Z score 4.11; also in the parietal lobe, precuneus, with peak at $(1, -37, 42)$, with a Z score 3.89;

[20] See Gusnard and Raichle (2001) for a recent illuminating discussion on the role and interpretation of the "baseline" conditions in brain imaging.

(3) a region in the occipital lobe, lingual gyrus, with peak at $(-15, -91, -14)$, with a Z score 4.1;
(4) a region in the left frontal lobe, superior frontal gyrus, with peak at $(-15, -13, 63)$, with a Z score 4.1.

The frontal and occipital activations have a weaker mirror image in the opposite hemisphere.

The region of activation in the difference PAC–RC are similar to the ones above. More specifically, the most active areas are:

(1) a region in the frontal lobe, lower than that observed in PAC–AC, with a peak at coordinates $(46, 39, -9)$, with a Z score 4.46;
(2) a region in the occipital lobe with a peak at $(-10, -91, -14)$, with a Z score 4.37;
(3) a region in the parietal lobe, precuneus, with a peak at $(15, -42, 50)$, with a Z score 4.11.

In contrast, it is clear from the tables for the AC–RC and RC–AC that there is little differential activation in these two cases.

In summary, the PAC condition provides qualitatively different activation than the AC and RC conditions. This finding stands in surprising contrast with the reasonable idea that a partially ambiguous lottery is an intermediate state between a totally ambiguous and a risky lottery. But it is consistent with the idea that the PAC condition is a less familiar experience for our subjects.

4.2.4. Frontal areas

There seems to be no strong activation of the higher frontal regions. More precisely, there is no difference displaying a strong and significant level of frontal activation in the levels above $z = 11$ mm. With one exception that we discuss later, this is also true in the differences PAC–AC and PAC–RC. In the first case, the frontal activation we have already reported is in the z interval between $+11$ and -11 mm. The same area is found in the difference RC–AC, but not in the PAC–RC difference.

The mentioned partial exception in the PAC–AC treatment is the region in the superior frontal gyrus of the left frontal lobe reported earlier (peak at $(-15, -13, 63)$). A similar activation is in the PAC–RC difference. In this case the peak is at $(-12, -8, 61)$, in the medial frontal gyrus of the left frontal lobe. The z coordinate is -4.7 mm, the highest in the PAC–RC difference.

It is worth observing that the pre-frontal cortex (PFC)[21] does appear prominently among the regions that are activated. The PFC is associated with planning, namely the ability to organize cognitive behavior in time and space. [22]

[21] This is the pole of the frontal lobe. It corresponds to the Brodmann areas 9, 10, 11.

[22] This is by now a classic finding. It has first been suggested by lesion studies (see for example the early studies of Shallice, 1988). These early results have been confirmed by brain-imaging studies. See, e.g., Zalla et al. (2000), Koechlin et al. (1999, 2000). However, the literature on this is very large: a useful review is Cabeza and Nyberg (2000). Owen (1997) offers a detailed review of definition and properties of planning ability in human subjects.

4.2.5. Orbito frontal ventromedial areas

There seems to be no strong activation of the ventromedial sections of the frontal lobes. That is, of areas known to mediate the processing of somatic and emotional reactions. A partial exception is an area that appears the RR–PAR difference; the peak is at $(6, 19, -18)$, right cerebrum, frontal lobe, medial frontal gyrus. The score is $Z = -3.47$, $p < 0.00027$. This is the only significant exception: the relative activations in AC–AR at $(1, 32, -22)$ and in RC–RR at $(-1, 8, -18)$ are likely to be artifacts, since they are at the extreme outer boundary of the brain.

5. Conclusions

We conclude by first summarizing the findings of greatest significance (in Section 5.1), and then by providing these findings with a provisional interpretation (in Section 5.2).

5.1. Summary of findings

(1) In their choices, subjects behave as predicted by models of risk and ambiguity aversion; their ambiguity aversion is consistent across the PAC and AC conditions.
(2) The time to decide is shorter in the C conditions. Among those, the minimum is attained in the PAC and AC conditions.
(3) Learning seems to occur in the PAC condition, less so in the other C conditions, and is almost absent in the R conditions.
(4) In the PAC condition, a larger distance from the cutoff point of the certain value makes the decision faster.
(5) The regions with most intense activation are observed in the C conditions, particularly in the difference between PAC and AC.
(6) There is low activation of ventromedial regions.
(7) There is low activation of the high frontal and pre-frontal regions.
(8) The only important frontal activation is in the PAC condition.
(9) There is high activation in the parietal regions in the C conditions.

5.2. Interpretations

5.2.1. A possible choice procedure

The results we have reported strongly suggest that a computational model of decision making might give a more accurate model of the behavior of decision makers.[23]

Here is procedure which gives an account of the observed behavior of subjects in the C condition. In all three cases (whether the main lottery is R, or PA, or A), the subjects are comparing the certain value with some estimate of the value of the main lottery. When this lottery is R the estimate is in substance a sum of the two outcomes, perhaps followed by a simple division. In the A case, the subject considers the best and worst possible scenarios.

[23] A similar idea is developed in Dickhaut et al. (2003).

In the best scenario, all balls are of the "good" color (the one that gives the largest payoff), and in the worst scenario they are all of the "bad" color. In both cases, the corresponding lottery is degenerate, yielding with certainty the prize associated with the only type of ball existing in the urn. So it is easy to evaluate. The situation is more difficult in the PA condition. In this case, the same process of reduction to the best and worst case yields two non-degenerate (true) lotteries: one yielding the high prize with probability $1/18$, the other with probability $17/18$. The subjects then use some rule that takes into account this best and worst case evaluations to estimate the main lottery.[24] Notice that according to this procedure, the PA condition is the hardest of the C conditions, and it is not intermediate between the other two, as it may appear from other perspectives.[25]

We note that in our experiment the choices are similar to those predicted by economic theory. This is not necessarily going to happen in general. We expect that as decisions become more complex, the constraints on the procedure delivering choices will become increasingly important, and affect in a systematic way the decision itself.

The procedure we have outlined may not be consciously followed by the subjects. There is however a substantial difference in the response time in our experiment (always less than three seconds) and that observed in simple computational problems (for example in the cited studies by Dehaene and co-authors). This difference suggests that the procedure in our study does not involve explicit calculations and may be partially automatic. We consider this issue important because automatic processes need not be mediated by consciousness. As a consequence, they are likely to produce relatively inflexible behavior that differs from the repertoire produced by conscious or planned thought. Clearly, more research is needed in this arena.

5.2.2. Approximate and exact estimates

The evidence we have presented suggests that subjects develop their decision process trying to provide some quantitative estimate of the lotteries, but that these estimates are approximate rather than exact. This conclusion is suggested first by the short response time, particularly short in the harder tasks, and it is supported by the observation that the computational aspects of the estimates used in the decision are located in the parietal, rather than frontal lobe.

This statement is significant and informative only inasmuch there is a qualitative difference between exact and approximate processes. For example, a difference in the cerebral networks activated by the two types of processing. This is precisely the conclusion reached by a set of recent studies by Dehaene and different co-authors (see Dehaene et al., 1996, 1999; also see Pugh et al., 1996 and Jonides et al., 1999). These studies argue for the exis-

[24] An example of one such rule is the so-called "α-MEU" rule, according to which an action is ranked via a convex combination of the best and worst case evaluations (α is the weight given to the "min" component, a measure of the subject's ambiguity aversion; see Ghirardato et al. (2002) for details and an axiomatic treatment of this rule).

[25] For instance, in terms of the amount of information available to the decision maker. There is only one possible composition of the urn in the risky lottery; in the partially ambiguous one, there is a set of possible compositions, and in the ambiguous one there is an even larger set. Or consider the point of view of a decision maker who is evaluating lotteries according to the MEU model. The worst case in the risky lottery is better than it is in the partially ambiguous one, and this is in turn better than it is in the ambiguous one.

tence of a specialization for processing *approximate* numerical quantities that is common to humans and animals, particularly mammals.[26] In addition, exact and approximate processing are associated with activity in different cerebral locations. For example in (Dehaene et al., 1999, p. 971) the authors note that

> "[...] the bilateral parietal lobe showed greater activation for approximation than for exact calculation. The active areas occupied the banks of the left and right intra parietal sulci, extending anteriorly to the depth of the post central sulcus and laterally into the inferior parietal lobule [...] Activation was also found during approximation in the right precuneus, left and right pre central sulci [...]"

These two regions also relate differently to language centers. In behavioral and brain-imaging studies exact calculations are shown to be language dependent, while approximations rely on a visuo-spatial cerebral network.[27]

As we have seen, the PAC condition has a comparatively strong activation of the frontal lobe, which extends over a large part of the middle frontal gyrus. This is one of the findings that sets the PAC condition apart from the others, including the AC and RC conditions. There are two possible interpretations of this difference. The first is that some exact calculation is taking place when subjects are considering a partially ambiguous lottery. This is partially in agreement with the finding of Dehaene et al. (1999), but is not entirely convincing in view of the short response time in this condition. A second interpretation appears more convincing on the basis of the evidence we have presented so far: more general higher cognitive functions become involved over the course of the trial, as subjects try to arrive at a satisfactory method to evaluate the PA lottery. Again, further research is necessary here.

---

[26] For instance, in (Dehaene et al., 1999) the authors state that

> "Within the domain of elementary arithmetic, current cognitive models postulate at least two representational formats for number: a language-based format is used to store tables of exact arithmetic, and a language-independent representation of number magnitude, akin to a mental 'number line.' "

See, e.g., Dehaene (1992) for a review of these findings.

[27] See Dehaene et al. (1999) and Spelke and Tsivkin (2001). In the Spelke and Tsivkin's (2001) study, subjects were familiar with the two languages (Russian and English). They were trained to execute mathematical tasks either approximately or exactly. The performance after training improved, so training was effective. However, the crucial test was the performance on new tests. When tested on the problems to be solved exactly, the performance was significantly better when the test was administered in the same language in which it had been taught, independently of whether it was English or Russian. On the contrary, the performance on approximate tests was independent of the language. In the authors' words:

> "[...] a specific, natural language contributes to the representation of large, exact numbers but not to the approximate number representation that humans share with other mammals. Language appears to play a role in learning about exact numbers in a variety of contexts [...]"

5.2.3. The response times

Let us recall the two facts that stand out. First, response times are longer in the R than in the C conditions. Second, among the C conditions, response time is shortest in the PAC condition. At first blush, these facts seem to directly contradict our procedural explanation's claim that the choices in the C conditions are harder, and that among them the PAC condition is the hardest. If this is the case, then why do the subjects not take more time in examining the more complex choices, and seem to do the opposite?

Consider this argument more carefully. It is based on two implicit assumptions: (1) that the allocation of attentional effort is in some way optimal, and (2) that subjects make at the beginning of the choice process a single decision on the amount of effort to be devoted to decision making.

The first assumption is reasonable, but its implications are richer than just that longer time will be used for harder problems as long as attentional effort is not costless. If it is costly, then the cost of such effort—which may be different in different conditions—will be compared to its effectiveness. The data on activation seem to indicate that the effort in the C conditions is more intense, hence perhaps more unpleasant. (Moreover, it is also possible that this effort is less effective that it is in the R condition.)

On the other hand, the second assumption is clearly false: Subjects monitor their own decision process, and they likely get a feedback on the effectiveness of their thinking process. This is a common assumption in models of attention (see, e.g., Bundesen, 1990, where attention produces a sharpening of the information, until the subject decides that it is optimal to decide).

Summarizing these considerations, it seems to us that in a realistic model of optimal allocation of attention, the time actually devoted to choice in hard conditions might actually be *shorter*, rather than longer. Thus, we do not think that the data on response times are necessarily at odds with the choice procedure suggested above. Moreover, it is worth remarking that the additional time in the R conditions *vis à vis* the C conditions might be also explained by the fact that in the C conditions only one of the two lotteries (the non-degenerate one) needs to be evaluated.

5.3. Decisions and emotions

5.3.1. The Somatic Marker hypothesis

The interpretation we have provided views the process of choice essentially as a cognitive process. This view contrasts with an interpretation suggested in the last decade—and based largely on neuro psychological and clinical observations—that "decision making is a process guided by emotions" (see, e.g., Bechara and Damasio, 2003).

This latter interpretation is centered around the Somatic Marker hypothesis (SMH; Damasio, 1994; Damasio et al., 1991). The initial insight for this hypothesis is provided by William James' theory of emotions. Let us first distinguish between *emotion* as the set of somatic reactions induced by an outside event (like the appearance of a snake) and *feeling* as the subjective perception of the events (both external to the subject and internal to the subject, as somatic reactions). In James' theory, the feeling is induced by the somatic reaction to outside events: in his beautiful expression, we are sad because we cry, we do

not cry because we are sad.[28] The SMH extends this idea to decision making. According to the hypothesis, "an emotional (that is, somatic) mechanism rapidly signals the prospective consequences of an action, and accordingly assists in the selection of an advantageous response option" (Bechara and Damasio, 2003).

The centers in the brain where such response is located are the orbito-frontal cortex and the amygdala. As we have seen, neither of these two centers is activated in our subjects. These two regions are routinely observed in studies conducted with the same scanner and techniques, so the failure to observe is almost certainly due to a lack of activation. [29]

The SMH has been tested in a standard laboratory experimental setup, in particular in the Card Deck Test (Bechara et al., 2000). In this test, subjects choose one deck of cards out of a set of four for a number of periods (usually one hundred). After they choose a deck they pick the top card, and a monetary amount associated to each card, which can be positive (gain) or negative (loss), is revealed. Normal subjects tend to choose, after an initial number of periods, decks that have positive expected return, even if the positive amounts are smaller. Patients with lesions in the orbito frontal region or in the amygdala tend to choose, even in later periods, deck that have larger positive amounts but compensated by even larger negative returns, so that the expected return for the deck is negative (Bechara and Damasio, 2003, Fig. 4). These results support the hypothesis that the orbito frontal cortex and amygdala are involved in decision processes, and they have been confirmed by imaging studies (see, e.g., O'Doherty et al., 2001).

There are several differences between the two experimental setups that can explain this difference. First of all, our subjects do not receive any information on the consequences of their choices during the experiment (no feedback), so they do not experience gain or losses. Second, our subjects do not learn anything about the distribution of the outcomes in addition to what they know at the beginning of the experiment. That is, we study specifically choice, rather than learning and choice. In contrast in the Card Deck Test subjects experience incremental learning: they are informed of the outcome in each period, and can use this information in the following choices. Finally, our subjects have only gains, while subjects in the Card Deck test have gains and losses. It is interesting to note that in a study (Dickhaut et al., 2003) with a structure similar to the present study, but involving losses, an orbito frontal activation appears in the comparison between gain and losses.

[28] More precisely:

> "[. . .] the more rational statement is that we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble, because we are sorry, angry, or fearful, as the case may be" (James, 1884).

[29] For example an orbito frontal activation appears clearly in the study (Dickhaut et al., 2003), conducted with the same devices and techniques. In the images of the present study, a clear example of a ventromedial orbito frontal activation can be found in any of the subtractions from ECR of any of the conditions. A particularly clear instance is for example in the ECR–AR subtraction, between the vertical coordinates 0 to −13 (see the set of images ECR in http://www.econ.umn.edu/~arust/neuroecon.html). A comparative activation of this region in the ECR condition is a standard—although not yet well understood—finding.

5.3.2. Reward anticipation and outcomes

Very similar differences appear between the setup of our study and recent studies focusing on the neuro anatomical and neuro chemical mechanisms underlying the evaluation of rewards, and in particular on the separation between expectancy and experience of reward and loss (see Breiter et al., 2001, but also Breiter et al., 1997), or reward anticipation and the reward outcome (see Knutson et al., 2001).

In these studies, the experimental sessions consist of a sequence of trials, where subject observe a cue that may signal the delivery of a reward or the lack of reward (or of a loss). Immediately after the arrival of the cue and possibly of their response, subjects know whether a reward is given in that trial. Different regions are activated in the different cases (with Nucleus Accumbens, Ventromedial Frontal Cortex and Orbito Frontal Cortex the main regions in the various cases). These studies provide the foundation for a mechanistic explanation of the processes that go from delivery of reward or lack thereof, to subject's evaluation of the outcome. However, they are not a study of the process leading to choice.

Further studies, with a careful analysis of brain activation, are needed to test the hypothesis that choice separated from immediate reward requires different brain processes, and to understand how emotions enter the decision process. It seems, though, that the black box of the decision process is slowly beginning to yield.

Acknowledgments

We thank a large number of colleagues that have listened to formal and informal presentations of these results, for comments and insights; in particular, the participants of the Minnesota Conference on Neuroscience and Economics. This research was financed in part by the NSF Grant 0136556 to Aldo Rustichini.

Appendix A. Regression tables

Table A.1
Regression on response times for the C conditions

Variable	Coeff.	Std. error	t	$P > t$	95% conf. interval
Response time for AC on discut and order					
Discut	−34.68	7.77	−4.46	0.000	[−50.03, −19.32]
Order	−3.91	12.68	−0.31	0.758	[−28.95, 21.12]
Constant	2411.67	165.70	14.55	0.000	[2084.53, 2738.81]
Response time for PAC on discut and order					
Discut	−56.71	8.44	−6.71	0.000	[−73.39, −40.03]
Order	−20.99	13.46	−1.56	0.121	[−47.58, 5.59]
Constant	3106.25	172.11	18.05	0.000	[2790.70, 3417.4]
Response time for RC on discut and order					
Discut	−53.61	12.61	−4.25	0.000	[−78.53, −28.69]
Order	−29.35	16.28	−1.80	0.074	[−61.53, 2.82]
Constant	3286.75	188.70	17.42	0.000	[2913.89, 3659.62]

Table A.2
Regression on response times for the R conditions

Variable	Coeff.	Std. error	t	$P > t$	95% conf. interval
Response time for AR on value, order and variance					
Value	4.24	12.04	0.35	0.725	[−19.58, 28.07]
Order	−11.01	20.51	−0.54	0.592	[−51.60, 29.58]
Variance	219.26	108.28	2.02	0.045	[5.005, 433.51]
Constant	2470.72	890.28	2.78	0.006	[709.15, 4232.29]
Response time for PAR on value, order and variance					
Value	21.98	12.89	1.700	0.091	[−3.53, 47.49]
Order	−1.21	22.02	−0.06	0.956	[−44.78, 42.36]
Variance	311.23	117.17	2.66	0.009	[79.42, 543.04]
Constant	1333.62	917.55	1.457	0.147	[−481.52, 3148.77]
Response time for RR on value, order and variance					
Value	12.46	13.22	0.94	0.348	[−13.69, 38.62]
Order	−8.48	22.70	−0.37	0.709	[−53.41, 36.44]
Variance	251.10	120.21	2.09	0.039	[13.26, 488.94]
Constant	1915.64	968.55	1.98	0.050	[−0.667, 3831.95]

Appendix B. The lotteries

All lotteries were built on the basis of an urn containing 180 balls, that could be either red or blue. The different lotteries were described by different proportions of red and blue balls in the urn, different information and different value associated with each ball. In all treatments, the color of the ball with the high value outcome changed over the different choices in that treatment.

B.1. Reference lotteries

The certain lottery C was a degenerate lottery: a single value would appear on the screen, ranging from a minimum of 10 to a maximum of 50.[30]

In the risky lottery R the urn had 90 blue and 90 red balls; the outcomes had expected values ranging from 30 to 40.[31] For each of the different expected values, we had three different lotteries, with different variance. For instance for the expected value 40 we had (80, 0), (58, 12) and (48, 32) as possible outcomes.

B.2. Main lotteries

In the R lottery, the urn had 90 blue and 90 red balls, and the monetary payoffs were fixed to be (60, 10).

In the partially ambiguous lotteries (PA), the urn contained at least 10 balls of each color, while the others could be of either color. In the ambiguous lotteries (A), *all* the balls could be of either color.

[30] Precisely: 10, 15, 20, 25, 28, 30, 31, 32, 33, 34, 35, 36, 38, 40, 42, 45, 50.

[31] Precisely: 30, 32, 35, 38, 40.

In the PAC condition, the payoffs were fixed to be $(60, 10)$; only the attribution to one or the other color was changed across lotteries. The same values, $(60, 10)$ were used for the A lotteries in the AC condition. In the PAR and AR conditions, the PA and A lotteries had a simple outcome structure: five different pairs of outcomes. One was always equal to zero, the other ranged from 60 to 80.[32]

In the implementation of the lottery in the final stage of the experiment (when the payment to the subjects was decided), the actual composition of the urns used for the ambiguous and partially ambiguous lotteries were drawn from a uniform distribution over the number of blue balls.

Appendix C. PET

PET measures the amount of *regional Cerebral Blood Flow* (rCBF) to specific regions of the brain. The procedure begins with the slow injection of a lightly radioactive liquid into an arm vein. The scanning begins almost immediately after the injection.

C.1. What PET detects

In a PET study, a subject is administered by injection a radioisotope emitting positrons (positively charged electrons). The isotope then circulates through the bloodstream to reach, among others, the brain tissue. Positrons are positively charged electrons, emitted from the nucleus of radioisotopes that are unstable because they have an excessive number of protons and a positive charge. When a positron comes in contact with an electron, the two particles annihilate turning the mass of the two particles into two gamma rays that are emitted at 180-degree to each other. These gamma rays easily escape from the human body and can be recorded by external detectors. The tomography detects these coincident rays, which indicates that positron annihilation has occurred somewhere along that coincidence line. The scanner then reports the amount of radiation from all different positions in the brain on average over the period in which the scan is taken. When the gamma rays interact with scintillation crystals, they are converted into light photons in the crystals. The scintillation events can be compared among all opposing detectors along many coincidence lines.

The procedure is reliable, accurate, and gives a complete picture of the brain, with a uniform precision for deep and superficial structures. However, it is necessary to take averages of rCBF over a relatively long period (on the time scale of the experiment) and the technique is therefore not suitable to detect changes that take place in short time intervals. (See, e.g., Phelps, 1992 for details.)

C.2. Method in our study

In our study, PET was used together with a tracer (H215O) to estimate rCBF, a standard indicator for brain activity. The rCBF was estimated from tissue radioactivity (after

[32] Precisely: 60, 64, 70, 76, 80.

correction with measured two-dimensional attenuation) using a Siemens ECAT 953B scanner (Knoxville, TN USA) with septae retracted; i.e., three-dimensional acquisition. An arm vein was used for access. The participant's head position was stabilized with a vacuum-molded pillow. A slow-bolus of H215O was injected intravenously (9.25 MBA or 0.25 mCi/KGB initially, infused at a constant speed over 30 s). Data acquisition (correcting for random decay and electronic dead time only) commenced upon arrival of activity into the head as evidenced by consistently rising true counts. Each experimental scan of 90 seconds contained data from one type of lottery, e.g., CGS or RG. The interval between scans was about 10 minutes. Images were reconstructed by filtered back projection including non-orthogonal angles to a final image resolution of 10 mm full-width at half-maximum.

C.3. Statistical analysis

An exposition of the conceptual and statistical foundations of the analysis is given in Frackowiak et al. (1997). For each individual and each treatment, we have a four dimensional vector (x, y, z, rCBF) recording the rCBF at the location described by the (x, y, z) coordinates.

C.3.1. Normalization

The data are for each individual subject, with brains of possibly different size and shape. The data are normalized onto a standard brain, so that a point in the standard brain corresponds to the same point in different brains.

We then analyze each pair of treatments separately, subtracting at each voxel the rCBF of the two activations, and then subtracting from this number, one for each subject, the average over subjects. A two-sided test gives the probability that the difference is larger than zero under the null hypothesis that the treatment is not influential.

References

Bechara, A., Damasio, A.R., Damasio, H., 2000. Emotion, decision making and the orbito frontal cortex. Cerebral Cortex 10, 295–307.

Bechara, A., Damasio, A.R., 2003. The Somatic Marker hypothesis: A neural theory of economic decisions. Games Econ. Behav. Special issue on Neuroscience and Economics.

Breiter, H., Aharon, I., Kahneman, D., Dale, A., Shizgal, P., 2001. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron 30 (2), 619–639.

Breiter, H., Gollub, R., Weisskopf, R., Kennedy, D., Makris, N., Berke, J., Goodman, J., Kantor, H., Gastfriend, D., Riordan, J., Mathew, R., Rosen, B., Hyman, S., 1997. Acute effects of cocaine oh human brain activity and emotion. Neuron 19 (3), 591–611.

Bundesen, C., 1990. A theory of visual attention. Psychol. Rev. 97, 523–527.

Cabeza, R., Nyberg, L., 2000. Imaging cognition II: An empirical review of 275 PET and fMRI studies. J. Cognitive Neurosci. 12 (1), 1–47.

Damasio, A.R., 1994. Descartes' Error: Emotions, Reason and the Human Brain. Grosset Putnam, New York.

Damasio, A.R., Tranel, D., Damasio, H., 1991. Somatic Marker and the guidance of behavior: Theory and preliminary testing. In: Levin, H.S., Eisenberg, M.H., Benton, A.L. (Eds.), Frontal Lobe Function and Dysfunction. Oxford Univ. Press, New York.

Dehaene, S., 1992. Varieties of numerical abilities. Cognition 44 (1–2), 1–42.

Dehaene, S., Dehaene-Lambertz, G., Cohen, L., 1998. Abstract representations of numbers in the animal and human brain. Trends Neurosci. 21 (8), 355–361.
Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., Tsivkin, S., 1999. Sources of mathematical thinking: Behavioral and brain-imaging evidence. Science 284 (7), 970–974.
Dehaene, S., Tzourio, N., Frak, V., et al., 1996. Cerebral activations during number multiplication and comparison: A PET study. Neuropsychologia 34, 1097–1106.
Dickhaut, J., McCabe, K., Nagode, J., Rustichini, A., Smith, K., Pardo, J., 2003. The impact of the certainty context on the process of choice. Proc. Nat. Acad. Sci. 100 (6), 3536–3541.
Ellsberg, D., 1961. Risk, ambiguity, and the Savage axioms. Quart. J. Econ. 75, 643–669.
Ellsberg, D., 1963. Risk, ambiguity and the Savage axioms: Reply. Quart. J. Econ. 77 (2), 336–342.
Frackowiak, R., Friston, K., Frith, C., Dolan, R., Mazziotta, J., 1997. Human Brain Function. Academic Press.
Ghirardato, P., Maccheroni, F., Marinacci, M., 2002. Differentiating ambiguity and ambiguity attitude. J. Econ. Theory 118, 133–173.
Gilboa, I., Schmeidler, D., 1989. Maximin expected utility with a non-unique prior. J. Math. Econ. 18, 141–153.
Gusnard, D., Raichle, M., 2001. Searching for a baseline: Functional imaging and the resting human brain. Nature Rev. Neurosci. 2, 685–694.
James, W., 1884. What is an emotion? Mind 9, 188–205.
Jonides, J., Schumaker, E., Smith, E., et al., 1999. The role of parietal cortex in verbal working memory. J. Neurosci. 18, 5026–5034.
Knight, F.H., 1921. Risk, Uncertainty and Profit. Houghton Mifflin, Boston.
Koechlin, E., Basso, G., Pietrini, P., Panzer, S., Grafman, J., 1999. The role of the anterior frontal cortex in complex goal-directed behavior. Nature 399, 148–151.
Koechlin, E., Corrado, G., Pietrini, P., Grafman, J., 2000. Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. Proc. Nat. Acad. Sci. 97, 7651–7656.
Knutson, B., Fong, G., Adams, C., Varner, J., Hommer, D., 2001. Dissociation of reward anticipation and outcome with event-related fMRI. Brain Imaging 12 (17), 3683–3687.
Luce, R.D., 2000. Utility of Gains and Losses. Lawrence Elbaum Associates, London, UK and Mahwah, NJ.
O'Doherty, J., Kringelbach, M., Rolls, E.T., Hornak, J., Andrews, C., 2001. Abstract reward and punishment representations in the human orbito frontal cortex. Nature Neurosci. 4 (1), 95–102.
Owen, A.M., 1997. Cognitive planning in humans: Neuropsychological, neuroanatomical and neuropharmacological perspectives. Progress Neurobiol. 53, 431–450.
Phelps, M.E., 1992. Positron emission tomography. In: Mazziotta, J., Gilman, S. (Eds.), Clinical Brain Imaging: Principles and Applications. Davis.
Pinel, P., Dehaene, S., Riviere, D., LeBihan, D., 2001. Modulation of parietal activation by semantic distance in a number comparison task. Neuroimage 14 (5), 1013–1026.
Pugh, K., Shaywitz, S., et al., 1996. Auditory selective attention: An fMRI study. Neuroimage 4, 159–173.
Ramsey, F.P., 1926. Truth and probability. In: Braithwaite, R.B. (Ed.), The Foundations of Mathematics and Other Logical Essays. Harcourt Brace, New York.
Savage, L.J., 1954. The Foundations of Statistics. Wiley, New York.
Shallice, T., 1988. From Neuro-Psychology to Mental Structure. Cambridge Univ. Press, Cambridge, UK.
Spelke, E., Tsivkin, S., 2001. Language and number: A bilingual training study. Cognition, 45–88.
Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme, New York.
Zalla, T., Koechlin, E., Pietrini, P., Basso, G., Aquino, P., Sirigu, A., Grafman, J., 2000. Differential amygdala responses to winning and losing: A functional magnetic resonance imaging study in humans. Europ. J. Neurosci. 12, 1764–1770.

REMOTE COMMAND AND CONTROL COMPROMISES SOLDIERS' TRUST IN THEIR LEADERS

Martin Liberg
The University of Skövde
Skövde, Sweden

Kip Smith
Linköping Institute of Technology
Linköping, Sweden

We report an experiment on the effect of remote command and control on soldier performance and trust. The experiment was conducted with active duty soldiers and officers as participants at the military training camp at Kvarn, Sweden. Soldiers ran our paintball assault-lane twice, once with the officer present in the lane and once with the officer out of harm's way. Two sets of data were recorded, response times to the command to "Move!" and questionnaires on the soldier's trust in the leader. Trust was significantly greater and response times were significantly faster in the leader-present condition. The best-fit linear regression function reveals a significant negative association between the two data sets. We conclude from this result that (1) remote command and control is associated with a decrement in soldiers' trust in their leader and that (2) this decrement in trust is associated with compromised soldier performance.

INTRODUCTION

Network-centric defense

The domain of application of our study is a proposed redesign of the organizational structure of platoons in the dismounted infantry. The U. S. Army and the Swedish Ministry of Defense have independently started processes that will change how platoons are formed and how they are led (Brehmer & Sundin, 2000; Friman, 2000; Hébert, 2003). The proposed redesign is called network-centric defense in both Sweden and the US. Soldiers on the battlefield may no longer take their orders from an officer whom they can see. Instead, orders will be given by a leader located in a safe position away from the battlefield.

One consequence of this fundamental change in team dynamics is that the members of a platoon may or may not know their leader. In either case, the leader will be remote. A second consequence is that all communication between soldiers and officers must be mediated by radios or other portable communication devices. Accordingly, part of the vision for the infantry is to require all soldiers to carry various types of sensors and telemetry equipment so that their remote leader can see and hear what they see and hear (Hébert, 2003).

Previous Work

In previous experiments conducted with civilian participants, we used our "paintball paradigm" to test the hypothesis that leader presence matters (Pangburn, Freund, Pangburn, & Smith, 2003; Smith & Liberg, 2005). Paintball allows us to simulate in an ethically defensible manner the hazards of live fire. We manipulated leader presence at two levels, one in which the leader was collocated with the participant playing the role of soldier and one in which the leader was in a remote, safe location. The collocated-leader condition emulates current and historic operations of the dismounted infantry. The remote-leader condition is intended to simulate remote command and control.

We have consistently found significant decrements in performance associated the remote condition: our civilian combatants were slower to respond to the leader's commands when the leader was not physically present in the assault lane. Response time, a key measure of performance in any command and control environment, suffers when the leader is not present. This finding supports the hypothesis that remote command and control compromises soldier performance. This result must be qualified with the following caveat: when civilians play the role of soldier.

Hypotheses

In this study we tested that hypothesis again with actual soldiers as participants and tested two additional hypotheses. Our second hypothesis was that remote command and control compromises soldiers' trust in their leader. To inform that test, we asked soldiers to complete a self-report questionnaire on the subjective levels of trust experienced in the collocated- and remote-leader conditions. This new data set also allowed us to assess directly a third hypothesis: there is a significant association between the inferred decrements in trust and performance. We expected slower responses and less trust in the remote-leader condition.

Further, we expected the association between trust and response time to be significant and negative.

METHOD

Participants

Three platoons of privates in the Swedish Army serving a voluntary term of service took part in the experiment, 52 men and 2 women. The soldiers' age ranged from 18 to 22 years. These young people are a sample taken directly from the population to which the results are intended to generalize.

Three officers in the Swedish Army, two majors and a captain, volunteered to participate in the role of leader. The officers were not known to any of the soldiers and were not in their direct chain of command. A platoon of 18 privates served as associates. Their responsibilities were to interact with the soldiers and run the experiment. They were the time-keepers, data recorders, and the sniper.

Experimental setting and apparatus

The experiment was conducted inside a 30 x 10 m hewn-timber barn on the training facility for the dismounted infantry at Kvarn, 25 km northeast of Linköping, Sweden. The assault lane, sketched in Figure 1, was delineated by five barriers constructed from surplus materials (1.0 m wide by 1.5 m high). The down-lane distance between barriers was 5 m; the cross-lane distance was 6 m. A small cup containing five paintballs was placed behind each barrier. At the far end of the lane stood a protective wall with two horizontal slots (0.5 m long by 0.2 m high) at shoulder height. The wall formed the sniper's hardened position. The slots allowed the sniper to shoot at (and be shot by) the soldier. Nine targets, standard-issue human-head targets (huvudmål), four red and five blue, were stapled randomly on the face of the sniper's position. The soldiers acting as time keepers stood in an observation area outside the barn where they could readily observe the soldier work his way up the lane.

Design

In the collocated-leader condition, the officer was present in the lane with the soldier. The officer was under fire and, on occasion, got in harm's way. In the remote-leader condition, the officer joined the time-keepers in the observation area, positioning himself so he could see but not be seen by the soldier. In both conditions, the officer communicated with the soldier using a radio. There were two types of orders, to move (advance to the next barrier in the lane) and to shoot (at the targets in front of the sniper's bunker).

Dependent measures

The behavioral variable was the time it took for the soldier to complete an order to "Move!" The time interval ended with the soldier came to a full, controlled stop behind the next barrier. The subjective variable was the reported level of trust in the leader. Trust was assessed using an eight item questionnaire constructed by the Swedish Defense Research Agency for the express purpose of assessing trust in leaders (Svenmark, 2005, personal communication). The instrument uses a 7 point Likert-type scale with anchors in the middle and on both ends. Half the questions were reverse-scored. The raw data is the average response to the eight questions. Swedish and English versions of the instrument are available from the authors upon request.

Procedure

The experiment was conducted in Swedish. All participants, both privates and officers, read and signed an informed consent form alerting them to the dangers of paintball impacts.

The soldier's initial position was behind the barrier furthest from the sniper. After loading his or her rifle with the five paintballs from the small cup located behind the barrier, the soldier used a voice-activated head-set to inform the officer that he or she had reloaded ("laddat"). On the officer's order to fire ("eld"), the solder attempted to shoot the "hostile" (red) targets without hitting the "friendly" (blue) targets. When out of ammunition, the soldier informed the officer that the rifle was empty ("slut"). On the officer's order to advance ("framåt"), the soldier moved to the next protective barrier on the other side of the lane. The cycle was repeated at each of the five barriers. The 25 paintballs, 5 at each barrier, were the soldier's only ammunition. In the collocated condition, the officer did not provide covering fire and advanced at a casual pace and only after the solder had reached the next position. Thus there was no opportunity for the soldier to learn from the officer how to advance.

The soldier ran the lane twice, once in the collocated condition and once in the remote condition. The order of conditions was counterbalanced. Immediately after finishing the second trial, the soldier completed the trust instrument twice, first answering for the first trial and then for the second trial.

RESULTS

The hypotheses tested by the experiment were (A) remote command and control is associated with a decrement in soldiers' self-reported trust in their leader, (B) remote command and control is associated with a decrement in soldiers' performance as assessed by response time to the command to "Move!", and (C) there is a significant inverse association between the decrements in trust and response time.

Questionnaire data

A two-factor mixed-effects repeated-measures ANOVA was conducted to assess the significance of the manipulation of leader presence and the possibility of sequence effects. The sequence of data acquisition (collocated-first, remote-first) was the blocking variable at

two levels. The two sets of answers to the trust instrument was the repeated measure. The raw data were the average of a soldier's responses to the eight questions in the instrument.

Figure 2 is a graph showing the self-report data across the manipulation of leader presence for the two sequences of trials. The plot reveals two main effects and no interaction between sequence and the manipulation of leader presence. The plot contains error bars (standard errors) to help explain the lack of a significant interaction. The ANOVA indicates that both main effects are significant: for sequence, $F(1, 52) = 10.38$, $p < .003$; for leader presence, $F(1, 52) = 27.995$, $p < .001$, $\hat{\eta} = .12$, power =.76. These results support the hypothesis that remote command and control is associated with a decrement in soldiers' self-reported trust in their leader. There is no indication in these data of a 'learning effect'.

Move data

A second two-factor mixed-effects repeated-measures ANOVA was conducted for the behavioral data. The raw data were the average of the soldier's five responses to the command to "Move!" Figure 3 is a graph showing the response time data. The plot reveals a main effect for leader presence and suggests there may be an interaction between sequence and leader presence. The ANOVA indicates that the main effects for leader presence is significant, $F(1, 52) = 11.489$, $p < .0013$, $\hat{\eta} = .04$, power =.32. While the test for the interaction was not significant, $F(1, 52) = 1.243$, $p > .20$, the direction of the interaction supports the notion that there might be an auspicious carry-over (learning ?) effect of initial leader presence. These results support the hypothesis that remote command and control is associated with slower performance.

Linear association between trust and response time

In Figure 4 we plot the difference between the soldier's two assessments of trust on the horizontal axis and the difference between the soldier's two move times on the vertical axis. Greater trust in the collocated-leader condition generates negative numbers on the horizontal axis. Faster times in the collocated-leader condition generate positive numbers on the vertical axis. Thus, our third hypothesis leads us to expect this cross-plot to have a downward sloping trend.

To test that hypothesis, we conducted a simple, linear regression analysis. The resulting model is: Difference in time = 0.40 - 0.30 * Difference in trust. A one-sided t-test for the slope parameter (β, -0.30) is significant, $t(52) = -2.312$, $p < .012$. Analysis of residuals indicates no deviation from normality and satisfactorily constant variance at all levels of the predictor variable (trust). While trust explains only 9% of the variance in the data (r^2 = .09), the model supports the hypothesis that there is a significant association between trust and performance, with soldiers trusting less and responding more slowly in the remote-leader condition.

DISCUSSION

The expected results were obtained for both sets of data and for their association. Thus, remote command and control appears to compromise soldiers' trust and their performance. As noted by one of our reviewers, this result will likely be controversial. If these results generalize beyond the context of this experiment, e.g., to a professional army, it would be prudent to evaluate further the consequences of plans to transition the dismounted infantry to remote command and control.

This study is the culmination of an extended program of testing and paradigm development directed at identifying and mitigating barriers to cognitive readiness on the electronic battlefield. We have identified remote command and control as one such barrier. In a previous paper we discussed the design and testing of a device that shows promise to mitigate the decrement in response time (Smith & Liberg, 2005). We are seeking funding to continue this line of work to ascertain how and why trust mediates cognitive performance on the simulated battlefield and in organizations at large and to test the design of tools that may mitigate its loss.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD 19-01-1-0621. Soldiers, officers, and facilities were graciously provided by Markstridskolan i Skövde. Equipment and logistic support was provided by Linköpings paintball och femkamp. Views expressed in this work are those of the authors and do not necessarily reflect official US or Swedish Army policy.

REFERENCES

Brehmer, B. & Sundin, C. (2000). Command and control in network-centric warfare. In C. Sundin & H. Friman, (Eds.), ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6 (pp. 45-56). Stockholm: The Swedish National Defense College.

Friman, H. (2000). Decision making in a ROLF staff environment. In C. Sundin & H. Friman, (Eds.), ROLF 2010 - The way ahead and the first step: Försvarshögskolans acta C6 (pp. 67-74). Stockholm: The Swedish National Defense College.

Hébert, R. (2003). Psychological science and the transformation of the military. American Psychological Society Observer, 16(9).

Pangburn, K., Freund, S., Pangburn, H. & Smith, K. (2003). Beyond Milgram: An experimental study of leader

presence. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society, 364-368.

Smith, K. & Liberg, M. (2005). Paintball 2: A study of remote-controlled tactile stimulation for the simulation of leader presence. Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society, 364 - 368.

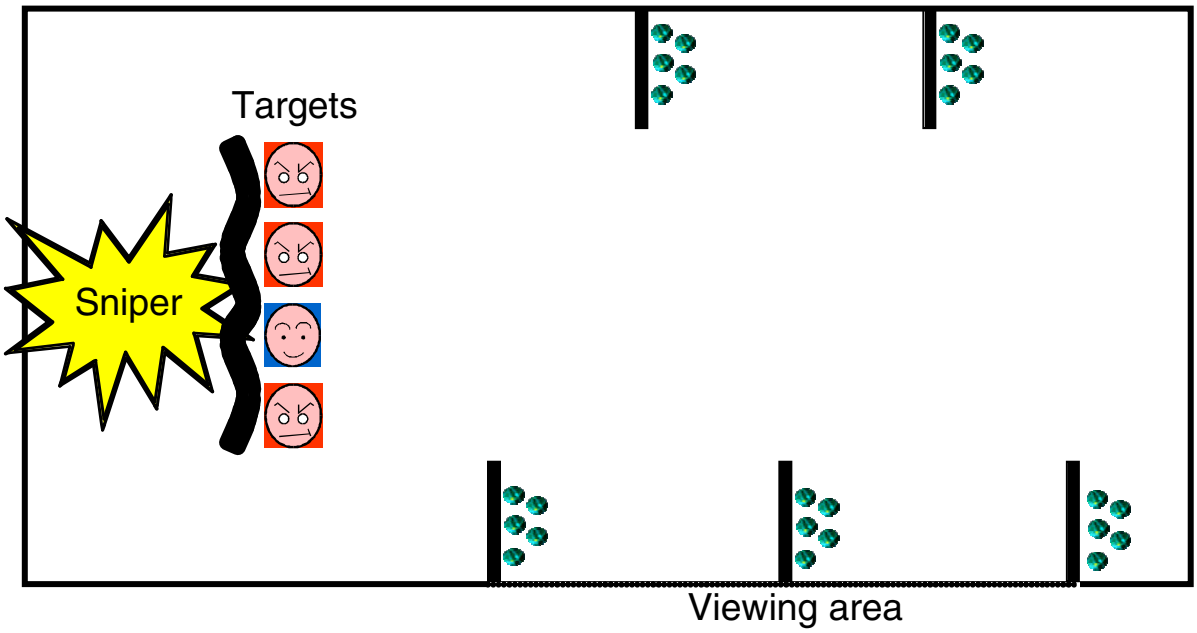


Figure 1. Layout of the paintball assault lane. Soldiers started behind the barrier closest at the bottom right and, on command, advanced from barrier to barrier toward the sniper to obtain ammunition with which they, on command, shot at the targets. In the remote condition, the leader stood hidden in the viewing area.

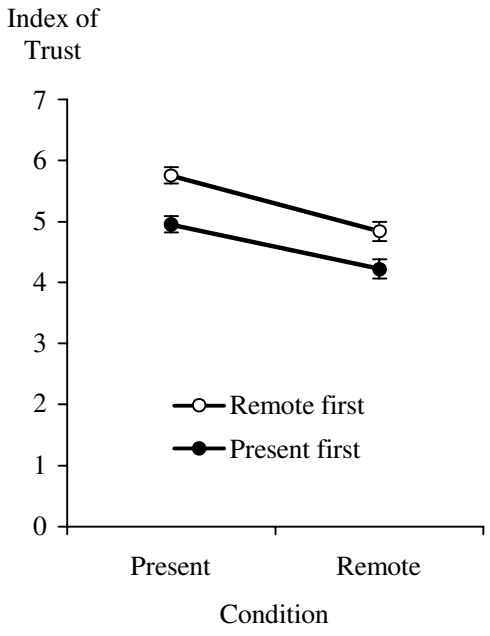


Figure 2. Means and standard errors of the soldiers' self-report data on the relative level of trust experienced in the leader-present and remote conditions showing higher levels of trust in the leader-present condition.

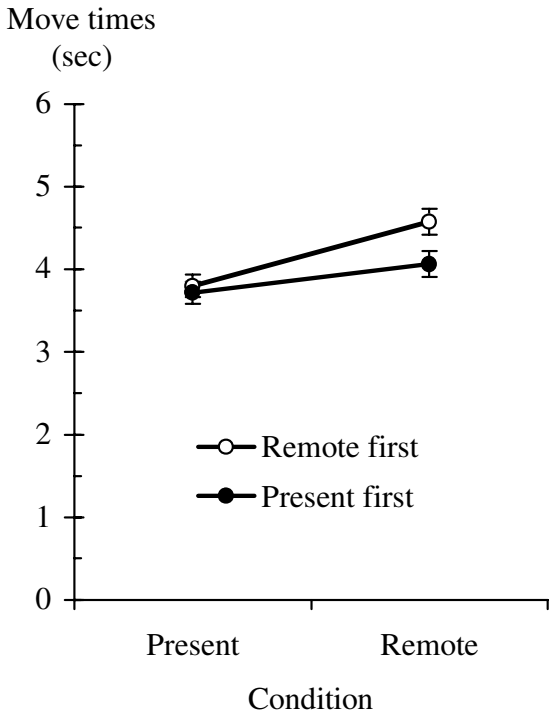


Figure 3. Means and standard errors of response times to the verbal command "MOVE!" showing faster responses showing faster responses in the leader-present condition.

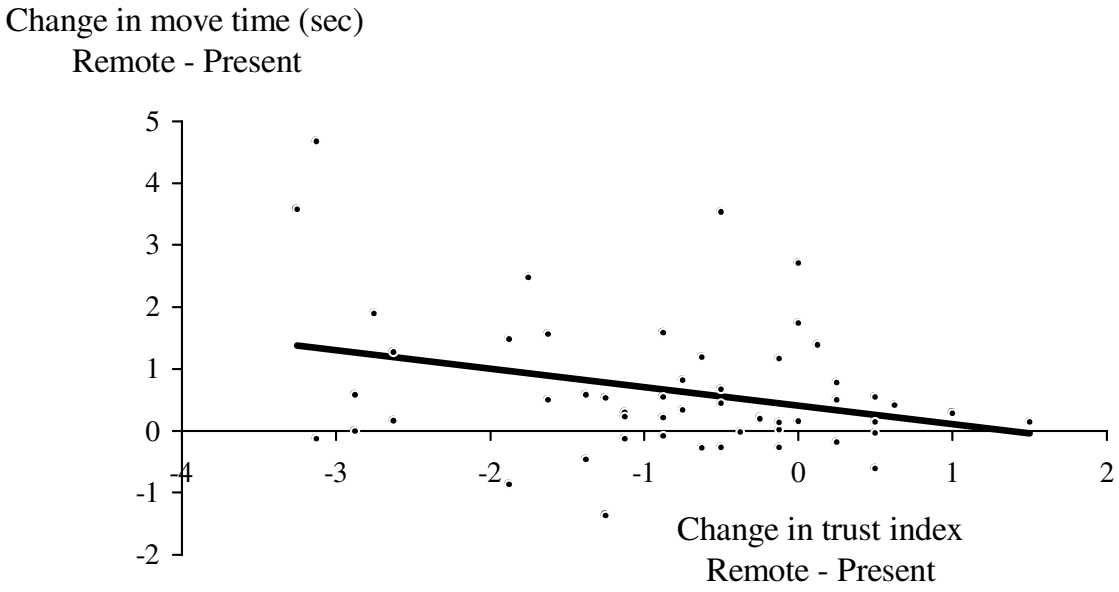


Figure 4. Scatter plot and best-fit linear regression function showing the significant negative association between the difference in experienced trust and the difference in move time.

THE 11TH INTERNATIONAL COMMAND AND CONTROL RESEARCH AND TECHNOLOGY SYMPOSIUM
Coalition Command and Control in the Networked Era

Cambridge
September 26-28, 2006

Using Microworlds to Understand Cultural Influences on Distributed Collaborative Decision Making in C2 Settings

Topics:
C2 Experimentation, Social Domain Issues, Cognitive Domain Issues

Authors:
Ida Lindgren [STUDENT] **and Kip Smith**

Contact information

Ida Lindgren
Ph.D. Student
Division of Industrial Ergonomics
Department of Mechanical Engineering
Linköping Institute of Technology
SE-58183 Linköping
Sweden
Phone: +46 13 28 1646
Fax: +46 13 28 2579
idali@ikp.liu.se

Kip Smith
Professor
Division of Industrial Ergonomics
Department of Mechanical Engineering
Linköping Institute of Technology
SE-58183 Linköping
Sweden
Phone: +46 13 28 27 64
Fax: +46 13 28 2579
kipsm@ikp.liu.se

Using Microworlds to Understand Cultural Influences on Distributed Collaborative Decision Making in C2 Settings

Ida Lindgren and Kip Smith

Linköping Institute of Technology

Abstract

As a means to facilitate coordination of international relief teams during sudden onset disasters, the UN has formed a structure called the On Site Operations Coordination Center (OSOCC). The main objective of the OSOCC is to coordinate international relief teams and help local authorities re-establish control in the affected area. As with any operation where people from different parts of the world are involved, multiculturalism can become an issue. Differences in values, norms and attitudes can create problems in communication, planning and execution of the operation. We use the C3Fire microworld and the Schwartz Value Survey as our main instruments to study cultural influences in command and control decision making in simulated OSOCC. The C3Fire microworld has been used extensively in research on networked-based command and control. Augmented with observation of a real OSOCC exercise, the experimental studies provide the basis for formulating clusters of behavioral differences in command and control that one can expect to encounter during an international operation. Results show that culturally-driven differences in planning and leadership style can pose potential barriers to efficient decision making in multicultural command-and-control centers.

Introduction

In 1988, a major earthquake struck Armenia. Many international relief teams came to help Armenian authorities to deal with the emergency. In the aftermath of the Armenian earthquake and other natural disasters, stories were told of international teams accidentally searching the same villages over and over while leaving other areas unsearched, resulting in many deaths. It became widely recognized that there was a need for more timely and coordinated response from international relief teams and that their work must not be a burden on the stricken country's resources. As a means to facilitate coordination of international relief teams during sudden onset disasters, a structure called the On Site Operations Coordination Center (OSOCC) was created.

The UN considers the OSOCC to be a coordination structure, not a command and control (C2) center. The purpose of an OSOCC is not to command international relief teams during the aftermath of a disaster. Rather, its purpose is to make impartial and transparent recommendations about how local authorities might choose to coordinate the efforts of the international relief teams. To use the label "command and control" to describe an OSOCC is therefore politically, and to some extent procedurally, incorrect. However, the recommendations made by the OSOCC strongly influence the relief effort and effectively function as executive directives. Moreover, the main objective of the OSOCC is to re-establish control in the affected area. Pragmatically, if not politically, it is appropriate to consider OSOCC work in the light of what is known about C2 decision making.

This paper has six parts. We first discuss culture and its potential impacts on decision making in C2 settings. The second part provides a description of the On-Site Operations Coordination Center (OSOCC) structure and function. The third part provides an overview of microworlds in general and of the C3Fire microworld specifically. The fourth part discusses the two studies we have conducted to capture cultural influences on decision making in C2 settings. The results from the two studies are presented in part five. The paper concludes with a discussion of the implications of the two studies for OSOCC specifically and command and control generally that proposes four clusters of culturally-driven differences in expectations and preferences for collaborative decision making.

Note that we have not conducted an exhaustive survey of how people from different cultures act in C2 settings. Our aim is more modest: to identify clusters of behavioral differences in C2 decision making that one can expect to encounter during an international operation. If clustering of differences becomes an accepted part of training programs for C2 personnel, it may help them bridge cultural barriers in OSOCC and other multicultural C2 settings.

Culture and C2 decision making

We argue that culture is a strong influential factor in international C2 operations. It is therefore relevant to first define what culture is. Most people have some idea, drawn from their own culture's folk psychology, of what the concept 'culture' means. These ideas tend to fall short because culture is a complex phenomenon. Triandis (1996) points out that while there are many definitions of culture, there is wide agreement that the elements of culture are shared by those with a common language, within a specific historic period, and a contiguous geographic location. Among these elements are customs, values and religious beliefs. These and other elements provide standards for perceiving, believing, evaluating, communicating, and acting. Culture is not encoded in our DNA but is propagated by interpersonal relations within a given physical environment (Duranti, 1997). In line with Duranti, Kim and Markus (1999) argue that cultures are composites formed by the immediate contingencies of specific sociohistorical circumstances and of individual actions. Culture emerges and is sustained by social relations within highly specific contexts. For succinctness, we adopt Smith and Bond's (1999, p. 39) definition and interpret it through the lens provided by Triandis and Duranti: a culture is a relatively organized system of shared meanings.

Culture's relation to cognition and communication

Our cultural heritage influences how we think, speak and act, and cannot easily be ignored. The influence of culture on cognition is partially revealed by studying language and communication. Innumerable studies have compared communication patterns across cultures (see e.g. Di Luzio, Günthner & Orletti, 2001) and found cultural influences on communication. Not only do we have different languages, but also different communication styles. For example, turn-taking differs remarkably between cultures. At one extreme, Swedes are known to listen quietly to the speaker and wait for their turn to talk (Daun, 1998; 1999). Interest in what the speaker is saying is displayed through silent attention. This style is reflected in the expression "*att tala i munnen på någon annan*" (to speak in someone's mouth) which means to speak at the same time as someone else. Interrupting or speaking in someone's mouth is considered very impolite and is something Swedish children learn at an early age not to do. Swedes, amongst others (e.g. Finnish people, Navajos in North America), are widely known for being a quiet people and for appreciating silence and solitude. In contrast, many Mediterranean cultures (e.g. Spaniards, Bosnians, Italians) encourage lively discussion where turn taking is less organized than in Swedish conversations and where the listener often shows his/her interest through talking aloud together with the speaker. In these cultures, there is no comparable expression as the Swedish 'to speak in someone's mouth', because that is simply how they are comfortable talking. It is therefore not impolite; on the contrary, it can be impolite to listen quietly. A quiet listener signals boredom. This does not mean that Mediterranean people like to communicate more than Scandinavians. It merely means that these cultures have communication styles in which participation and intent is displayed very differently (Daun, 1998). This example illustrates how conflicts could arise in multicultural groups. Consider a group of Swedes and Spaniards asked to work together with no prior experience with the other culture. Initially, their different communication styles would likely initially pose barriers to efficient cooperation, since each might perceive the other's communication style as insulting or rude.

Culture and values

In 1980, Gert Hofstede published his book *Culture's Consequences* which discussed a study that is still considered impressively extensive (Smith and Bond, 1999). He managed to collect questionnaire

responses from more than 100,000 individuals from around the world. The questionnaires concerned various aspects of employees' work experience that could be tied to fundamental human values. From this material, Hofstede was able to make comparisons across countries. The study now includes more than 60 nations (Hofstede & Hofstede, 2005). Hofstede's aim – and indeed it remains the aim of most of cross-cultural psychology (Smith and Bond, 1999) – was to map cultures on the basis of variables that can be directly linked to fundamental human values.

Hofstede (1980) used factor analysis to extract core 'dimensions' of human values, and compared relative weights given to each dimension across countries. This effort generated four bipolar core dimensions and national rankings for the sampled countries along each dimension. The dimensions are explained in Table 1. Definitions are taken from Hofstede and Hofstede (2005) and Smith and Bond (1999, p. 45).

Table 1: Hofstede's four core dimensions

Dimension	Explanation
Power distance:	The extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally
Uncertainty avoidance:	The extent to which the members of a culture feel threatened by ambiguous or unknown situations
Individualism /Collectivism:	Individualism: One's identity is defined by personal choices and achievements. In contrast, Collectivism: One's identity is defined by the character of the collective groups to which one is more or less permanently attached.
Masculinity /Femininity:	Masculinity emphasizes achievement. Femininity emphasizes interpersonal harmony.

While informative, Hofstede's rankings are averages calculated from thousands of individuals. As averages, they reveal only the central tendency and conceal the variability of individual differences. The rankings falsely imply that national cultures are "unitary systems free from variation, conflict and dissent" (Smith and Bond, 1999, p. 42). There is, of course, considerable individual variability within all cultures (Hofstede, 2001). Nevertheless, Hofstede's dimensions and rankings remain valid indicators of national culture.

More recently, Shalom Schwartz (1992, 1994) has become a central figure in theoretically-grounded cross-cultural psychology. The widespread acceptance of Schwartz's instrument and the explanatory power of his configural model of "value types" have led us to adopt the Schwartz model and methodology.

Schwartz rooted his study of values in a framework of human evolutionary needs (Schwartz and Bilsky, 1987, 1990). He identified in the previous literature a set of 57 human values (e.g., creativity, curiosity, pleasure, wealth, and health) each of which had been noted in more than one culture. He developed a survey instrument that he and his colleagues have used to collect data from individuals from more than 50 national cultures. The survey asks respondents to rate each of the 57 values "As a guiding principle in my life," using the nine-point scale shown in Figure 1. The –1 is unusual but highly useful. It allows respondents to indicate a "negative" value – a value they seek to avoid expressing or promoting through their choices and behavior.

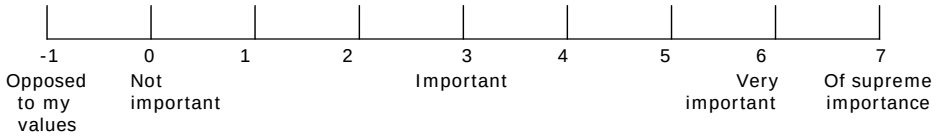


Figure 1: The 9 point response scale used in the Schwartz value survey. The extreme scores (-1 and 7) are used to anchor the ratings.

From the participants' responses to the 57 values, Schwartz constructed his configural model of 10 "value types", reproduced as Figure 2. The circular representation emphasizes the inter-relatedness of the 10 value types. Adjacent value types are proposed be to be most compatible and those on opposite sides of the circle to be in most conflict. The Schwartz model has been tested innumerable times since its initial publication. With few exceptions (e.g., certain regions in China), individuals in all literate cultures appear to implicitly distinguish the 10 value types when assessing the importance of specific values as guiding principles in their lives. The model represented by Figure 2 appears to be an exhaustive and near-universal classification of motivational values. Schwartz and others have used the instrument and the configural model of value types to explore and explain cross-cultural differences in a host of domains and applications.

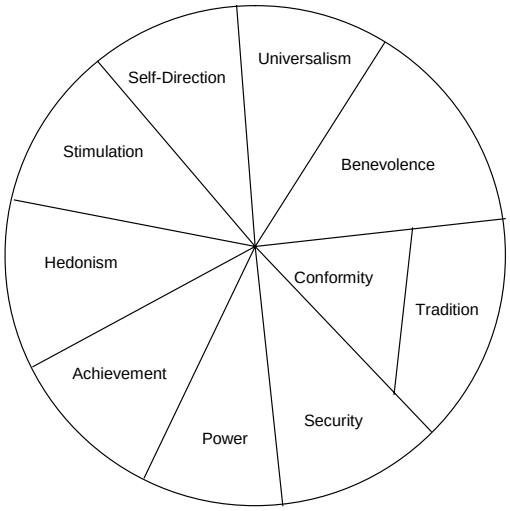


Figure 2: The configural model of the structure of core human values. (From Schwartz, 1992, 1994).

In contrast with Hofstede's dimensions, the Schwartz model uses the natural variability between individuals' answers as a source of explanatory power. His model also provides a basis for generating hypotheses that link responses to his value survey to performance measures (dependent variables), in our case captured by C3Fire. Different participants have different value structures that, ideally, correlate with differences in their patterns of behavior. By combining Schwartz's abstract level of measurement with our contextually specific measures from C3Fire, we can generate and test a myriad of hypotheses about cross-cultural differences in correlations between value types and performance variables in our simulated OSOCC.

In sum, people across different cultures differ in communication style and value priorities in life. The premise of our research is that these factors are likely to influence their decision-making styles and choice of strategies and tactics.

Culture and collaboration and decision making

Previous research has shown that there is more commonality than difference in decision-making style across cultures (Mann et al., 1998). However, differences do exist. Differences have been documented between Western and East Asian people (Chu, Spires and Sueyoshi, 1999) and between closely neighboring East Asian people (Chu, Spires, Farn and Sueyoshi, 2005). Mann et al (1998) investigated how decision making strategies differed across Western and East Asian samples and argue that what differs across cultures is a set of factors that determine (a) who makes the decision, and (b) the values and interests served by the decision. These factors include:

1) the authorities and entities invested with responsibility and control over decision making;
2) sources of expertise and advice;
3) whether it is an activity for the individual or the group;
4) the spheres in which individuals have freedom of choice; and
5) ideological principles and societal values that underlie decision rules and criteria for choice.

According to Mann et al., similarities and differences in roles, rights and responsibilities of individual decision making, all nested in the five points above, have received little attention in the cross-cultural literature. We are following Mann et al.'s lead to investigate how roles and responsibilities are distributed in culturally homogenous teams. Through investigation of cultural groups we are formulating clusters of differences in decision making that reflect the influence of cultural diversity.

OSOCC

The information about OSOCC presented in this section was obtained from (1) the United Nations' OCHA Orientation Handbook (2002), (2) the United Nations Disaster Assessment and Coordination UNDAC Field Handbook (2000), (3) OCHA's official homepage (http://ochaonline.un.org/), and (4) an interview with P. Becker, Head of Unit for Capacity at the Swedish Rescue Services Agency. Becker trains Rescue Service and other personnel for OSOCC work.

Background

When a disaster or emergency strikes a nation and there is an immediate need for coordination and support by international relief teams, the Office for the Coordination of Humanitarian Affairs (OCHA) can send members from the United Nations Disaster Assessment and Coordination team (UNDAC team) to the affected area. The UNDAC team is a stand-by team of disaster management professionals who are nominated and funded by member governments, OCHA, the United Nations Development Program and operational humanitarian United Nations agencies. The UNDAC team can be mobilized within hours of receiving a request from a country affected by an emergency or disaster. After natural disasters, such as earthquakes, the UNDAC team must be mobilized rapidly in order to effectively coordinate the search and rescue (SAR) operations of international SAR teams with local authorities.

While OCHA and its UNDAC team are responsible for ensuring that the humanitarian relief provided is effective, they are not responsible for providing that relief. The formal responsibility for all relief actions lies in the hands of the Local Emergency Management Authority (LEMA). The UNDAC plays a supporting role and assists the LEMA's efforts to coordinate the humanitarian relief effort. For this coordination to be efficient there can be one and only one coordination unit. If the local authorities elect to coordinate the work themselves without UN involvement, UNDAC will not become involved.

Coordination of international relief teams is not an easy task. The teams that arrive at the affected area talk different languages, have different backgrounds and training, and bring differing numbers of people and types of resources to the site. In fact, coordination is so difficult that UNDAC has created a coordination structure called the On Site Operations Coordination Center (OSOCC). The OSOCC is the actual physical location (or locations) where the UNDAC team does its work. The humanitarian community comes to an OSOCC to meet and exchange information and to get direction from the UNDAC team.

The OSOCC concept was introduced in the aftermath of earthquakes, but the structure and procedures behind the OSOCC make it a vital asset in any sudden onset disaster involving international relief resources. As a result it has been used with increasing frequency. There are no explicit or fixed criteria for when to set up an OSOCC. The guiding principles are: (a) the emergency must be a sudden onset disaster/emergency, where every lost hour can mean lost lives; (b) there is a need for coordination of international relief teams arriving to help local authorities; and (c) the local authorities need and have asked for support. The OSOCC is active during the first phase of relief operations and remains active

until the national/local authorities and/or the traditional UN relief structure can assume responsibility for coordination of the international relief effort.

The relationship between the UNDAC team and the local authorities (LEMA) can strongly constrain the activities of people working in the OSOCC. Because OSOCC personnel play a supporting role, their interactions with the LEMA must conform to local social conventions. Like all international personnel, the OSOCC personnel need to be sensitive to and follow local norms. Differences in norms for behavior, collaboration, and decision making can pose a challenge to the efficient coordination of humanitarian relief activities.

The OSOCC Concept

Figure 3 illustrates the OSOCC's role as a liaison between the OCHA/UN/LEMA and the international relief teams. OCHA is the governing UN office with headquarters in Geneva. The UN Resident Coordinator is the UN's representative in the affected country. During an emergency situation, he works in close cooperation with the OSOCC and is the UN official responsible for coordinating UN humanitarian assistance. He monitors and provides early warnings of potential emergency situations and by leading contingency planning to the UN. The LEMA is formally responsible for coordinating all relief operations.

The OSOCC has three parts: (1) the Reception/Departure Center; (2) the main OSOCC; and (3) possibly one or more sub-OSOCCs. The Reception Center is located at entry and exit points, such as airports or harbors. Its purpose is to assist airport authorities and to expedite the registration and arrival/departure of international relief providers. The main OSOCC is a physical location where decision makers meet. It is set up as close to the LEMA headquarters as possible. If needed, sub-OSOCCs are set up in remote locations. Sub-OSOCCs may operate semi-autonomously.

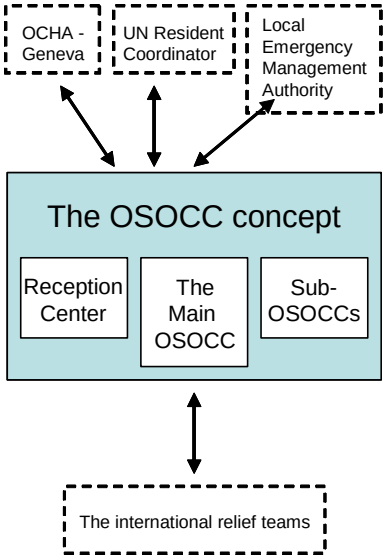


Figure 3: The OSOCC and its position within the overall structure of an UN-supported international relief effort.

The Main OSOCC

The main OSOCC, shown in the center of Figure 3, is a physical location manned by personnel from UNDAC and the international relief teams. Each international relief team is responsible for promoting the effective functioning of the OSOCC. To facilitate the efficient OSOCC operation, it is recommended that the international teams supply the OSOCC with manpower. Countries that support

large relief teams (e.g., Sweden) generally do so. It is difficult to say a priori how many people are part of a Main OSOCC. There are usually only four or five. There have been occasions when one UNDAC member was the entire Main OSOCC team. The international relief teams often manage to allocate some of their personnel to the OSOCC. Their participation helps ensure its efficiency.

The Main OSOCC has four goals. These goals are to (1) provide a system for coordinating and directing the activities of an international relief effort at the site of a disaster; (2) provide a framework for cooperation and coordination among the international humanitarian entities; (3) act as a link between such entities and the affected area's authorities; and (4) to coordinate the activities of international Search and Rescue teams as appropriate.

The cultural consequences of this ad-hoc and on-site team-formation can be daunting. The representatives of the various international teams generally do not know each other. They may or may not know members of the UNDAC team. They speak different languages. And yet they are charged with the task of working together immediately to coordinate a flood of humanitarian activity and to facilitate the LEMA's efforts to coordinate the relief effort.

Microworlds

Microworlds are simulated environments that realistically capture aspects of decision making problems (Johansson, 2005; Woltjer, 2005). With the use of microworlds we can bridge the gap between the confines of the traditional laboratory experiment and the "deep blue sea" of field research (Brehmer & Dörner, 1993). Brehmer and Dörner (1993) argue that microworlds (a) provide a task that can be made more complex, challenging, and realistic than traditional laboratory studies but that (b) generalize to interesting parts of real world problem solving while remaining (c) controllable and more easily analyzed than field studies.

Microworlds are characterized by the fact that they are *complex, dynamic* and *opaque* (Brehmer and Dörner, 1993). They are complex because they pose many goals, which means that the participants have to consider several options concurrently, e.g., different courses of actions or contradicting goals. Secondly, they are dynamic in the sense that participants have to consider different time-scales and unforeseen effects since the relationship between different variables are uncertain. Opaqueness refers to those dimensions of the simulation that are invisible to the participant, e.g., the speed, direction, and impact of wind. Opaqueness forces participants to make hypotheses and test them in order to understand and control the system. By enabling the experimenter to manipulate these dimensions, microworlds provide a much more tractable, reproducible, and flexibly designable research environment than a field study (Brehmer and Dörner, 1993; Gray, 2002). Furthermore, with microworlds the researcher is able to investigate questions that cannot be addressed in either field studies or laboratory experiments (Gray, 2002), and at the same time address issues that can be seen in real world tasks (Dörner and Schaub, 1994). It has been shown that experimental participants take well-designed microworlds seriously and become so engaged that their behaviour becomes completely natural and, accordingly, valuable to the researcher (Dörner and Schaub, 1994; Gray, 2002). These considerations make microworlds valid tools for studying the impact of culture on decision making in command and control settings.

Two studies designed to study cultural influences on decision making

We have conducted two studies of cultural influences on decision making in command and control in the OSOCC. The first study is a microworld study using C3Fire. The second study is an observational study conducted during an OSOCC exercise run by the Swedish Rescue Services Agency (SRSA) and the Danish Emergency Management Agency (DEMA). We first turn to the experimental study and its results and then continue with the observational study.

Study 1 – C3fire

OSOCC operations place strict constraints on the conduct of the experiment. OSOCC teams are formed ad-hoc and on-site. Team members may or may not know each other. Because there is no time for team-building, they get to know each other as they work. As they get to know each other, their way of working together is likely to evolve. These considerations led us to design the experiment to meet three sets of constraints: (a) we need to elicit and capture spontaneous but collaborative emergency-services decision making in response to a simulated emergency, (b) we need to emulate the ad-hoc nature of OSOCC team formation and to capture the actions for decision making that guide the development of teamwork in the simulated OSOCC, and (c) we need to gather individual self-report information about values.

Apparatus - The Microworld C3Fire

C3Fire (Granlund, 2002) is a microworld in which a group of people work together to extinguish a computer-simulated forest fire. The group's task is to collaborate in an experimentally controlled configuration for command and control interaction. The C3Fire microworld is distributed in a client-server configuration, meaning that each participant working in the simulation works at his own client PC. Their actions are logged in the C3Fire system and are observed by a researcher who manages the experiment.

The C3Fire microworld has been used extensively in previous research on network based command and control (Artman and Wearn, 1999; Granlund, 2002, 2003; Johansson et al., 2003; Woltjer, 2005), and comes form a long tradition of microworld research of distributed decision making (Brehmer, 1992, 2005; Brehmer and Dörner, 1993).

There are three classes of units in C3Fire; fire trucks, water trucks and fuel trucks. All units are interdependent. All four participants involved in C3Fire can control all units. Interdependencies among decision makers arise whenever different classes of fire-fighting units are assigned to different participants in the simulated OSOCC. For example, the locations and activities of water trucks and fuel trucks constrain the actions of the fire trucks. If different people have control over these different resources, their actions are mutually constraining. This provides ample opportunity for conflicts to arise.

The participants can communicate through e-mails only. We, as experimenters, did not establish an organizational structure for communication and control that the teams were to follow. E-mail communication and truck control were not constrained by the experimenters or the C3Fire software. As a result, all participants could (1) communicate with all other participants (they could send a message to one or all other participants) and (2) command all trucks (all fire, water, and fuel trucks) and, (3) override commands made by other participants. In short, all structure was left to the teams, much as it is in an OSOCC.

Every event in an experimental trial generates time-stamped data that C3Fire automatically records and stores. There are three classes of events: commands to trucks, the way the team allocates duties, and the content and structure of their communication. These are the major dependent variables gathered during the experimental trials. This paper concerns the commands to trucks only.

Participants

Fifty-five male participants who identify with three different national cultures, ages 19-37 (mean 24.8), participated in our experiment. Thirty-two of the participants identify themselves as Swedish, 8 as Iranian Swedes, and 15 as Indian. All participants signed an informed consent form. They were promised monetary compensation for completing two four-hour sessions of experimentation. All 55 participants completed the study and received their compensation.

The overarching goal is to identify clusters of expectations and behaviors for collaboration and organizational structure that (1) vary systemically across cultures and that (2) personnel in a newly-formed OSOCC can use to identify potential barriers to collaborative decision making. To this end it is

necessary to elicit expectations and spontaneous behaviors from a variety of cultures. Inspection of the pioneering work of Hofstede (1980), suggests that the Swedes, Iranians, and Indians differ strongly from one another along a variety of dimensions.

Procedure

The participants reported to the laboratory in groups of eight. In the laboratory, the eight were randomly and anonymously assigned to two teams of four OSOCC decision makers. The purpose of the random and anonymous assignment to teams was to minimize reputation effects and to emulate the ad hoc nature of OSOCC team formation. In what follows, the word 'team' signifies the four participants in a simulated OSOCC and is the basic unit of analysis in the experiment. The word 'group' is reserved for all eight participants when the two teams are brought together or to the larger ethnic group with which they identify. One Indian group consisted of seven participants and worked in teams of four and three.

The two teams worked in parallel in two different simulated OSOCC. This arrangement made it possible to gather data on two teams (two units of analysis) simultaneously. It also provided the opportunity to periodically reassign participants to teams to minimize reputation effects.

The cycle of activities

The procedure consisted of eight cycles, each including three sets of activities corresponding to the three constraints.

Activity 1 – C3Fire experimental trials

Each of the eight participants sat at a separate client computer and was linked to his teammates by the C3Fire software. Their only mode of communication was the email system provided by C3Fire. The two teams of four were connected to different server computers. The two servers independently ran the same C3Fire scenario concurrently.

Scenarios

We created eight different experimental scenarios. The teams encountered scenarios designated A through D on the first day of experimentation. They encountered scenarios E through H on the second day. Table 2 shows the systematic manipulation of three factors that generated the eight experimental scenarios. The factors are map, map rotation, and initial fire size. Two different maps (m1 and m2) with differing configurations of forests and houses, etc., form the foundation for the eight scenarios. Four scenarios use map m1 and four use map m2.

Table 2: Dependent variables and their manipulation in the C3Fire scenarios.

Scenario	Map	Map rotation	Fire size	Scenario	Map	Map rotation	Fire size
A	m1	0	2x2	E	m1	180	2x2
B	m2	0	2x2	F	m2	180	2x2
C	m1	90	3x3	G	m1	270	3x3
D	m2	90	3x3	H	m2	270	3x3

As shown in the third column of Table 2, map rotation was manipulated at four levels (0°, 90°, 180°, and 270°) to make the maps appear different. Initial fire size refers to the size of the fire, in squares, at the beginning of the scenario. This was manipulated at two levels: the larger the fire, the greater the challenge.

Anonymity

One of the salient social characteristics of an OSOCC is that its members may not know each other when they arrive on site. They get to know each other as they work. There is no straightforward way to capture this emergence from anonymity in an experimental setting in which participants may indeed know each other. We crossed this hurdle by bringing eight participants into the laboratory at once and

splitting them, randomly and anonymously, into two groups of four. Initially, no one on a team knew who the other three team members were. As they day advanced, they became better acquainted and had numerous opportunities to interact as a team.

Specifically, team membership was randomly assigned and unknown during the first C3Fire trial of each day. There was no opportunity for consultation before the trial. As in a new OSOCC, the team had little common ground (other than their shared ethnicity) but had to attack the emergency immediately. This situation was repeated for the second, fifth, and sixth C3Fire trial. Team membership was randomly shuffled and, once again, there was no opportunity for consultation. Participants repeatedly found themselves in a new OSOCC facing a new emergency.

Activity 2 – Group discussion

In the second part of each cycle, the team sat around a computer monitor, watched a replay of their C3Fire session, and engaged in an open-ended conversation about their session. During these 'after action reviews', most teams developed an organizational structure, allocated responsibilities, and debated alternative strategies for dealing with the emergencies posed by C3Fire. Their conversation was recorded using both a video camera and audio equipment for subsequent qualitative and quantitative analyses. The groups' discussions of the C3Fire sessions are not covered in this paper.

Activity 3 – Questionnaires

The participants filled out six different questionnaires designed to assess the participants' (1) demographic background, (2) personality traits according to the Big Five (NEO Five Factor Inventory, NEO-FFI: Costa & McCrae, 1992), (3) conflict avoidance, (4) tolerance for uncertainty, (5) time horizon, and (6) value priorities (the Schwartz Value Survey: Schwartz, 1992, 1994).

Results

Following Schwartz (1992), we have constructed plots of partial correlations between the 10 value types and (a) measures of fire-fighting effectiveness during the C3Fire trials and (b) the indices provided by the other survey instruments. Two of these plots are presented here. Figure 4 shows the patterns of partial correlations between the index of time horizon and the 10 value types. Strongly positive correlations reveal values that are associated with planning and a future-directed activity. Negative correlations reveal values that are associated with short-term goals. The correlations for the Indian participants are weak with one curious exception, the sharp downward spike at tradition. This single-point anomaly may be spurious. The overall flat pattern for the Indian participants suggests, for these individuals from this ethnic group, there is little association between values and planning for the future. Planning may not be a key activity for this group. In contrast, there are strong correlations, both positive and negative, between the Swedes' and the Iranians' rankings of the value types and the index of time horizon. Achievement, power, and security are positively correlated with planning for both groups. It appears that both of the groups raised in Sweden may perceive planning as a way to get ahead.

Figure 5 shows the patterns of partial correlations with the index of conflict avoidance. Strongly positive correlations reveal values that are associated with avoiding confrontation. Negative correlations reveal values that are associated with meeting challenges head-on. The difference between the Swedish and Iranian groups is striking. For the Iranians, both achievement and power are strongly and negatively correlated with conflict avoidance. Iranian participants who indicated a relatively high need to achieve are much less adverse to conflict than similarly ambitious Swedes.

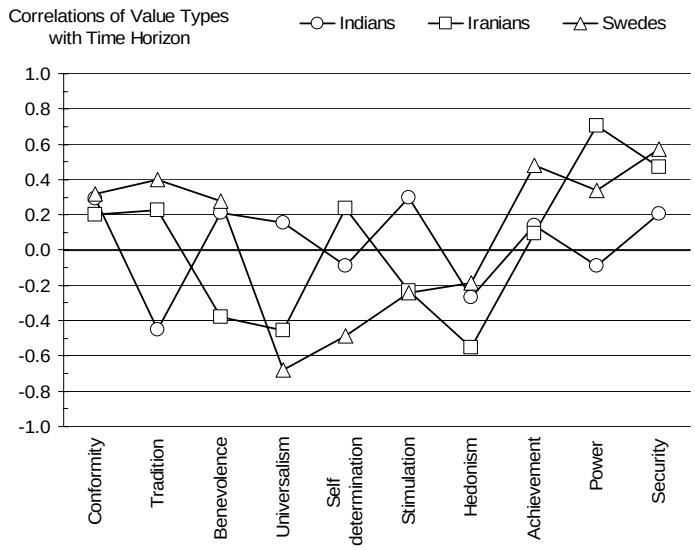


Figure 4: Associations between value types and the index of time horizon. Positive correlations indicate a common preference for the value type and actions that plan for the future.

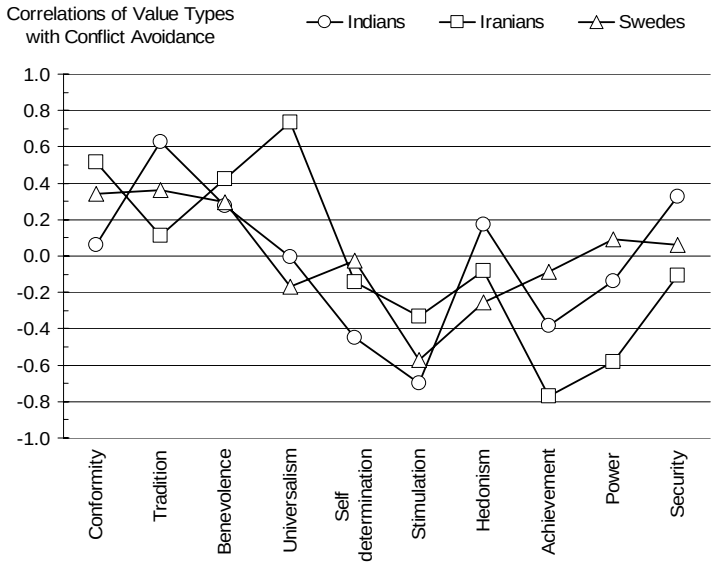


Figure 5: Associations between value types and the index of conflict avoidance. Positive correlations indicate a common preference for the value type and actions that seek to avoid conflict.

We have analyzed the allocation of responsibilities and distribution of trucks across the four participants (A, B, C and D) using plots of the relative frequency of commands they sent to the 12 trucks (F1, F2, F3, etc.). In these plots, a fully black cell represents the highest percentage of commands sent to a truck by a specific individual. At the other extreme, a purely white cell means that no commands were sent to the truck by that participant. Intermediate tones of grey represent intermediate percentages of messages in a linear mapping. Two cells that are equally dark therefore represent equal frequencies of commands. In Figure 6, we can see that participant A did not command any trucks, participant B sent commands only to gas trucks (G10-12), and participant C only to water trucks (W7-9). In contrast, participant D sent commands to almost all trucks, but concentrated on the fire trucks (F1-6). This distribution suggests that the team largely adhered to a strict partitioning of roles and responsibilities.

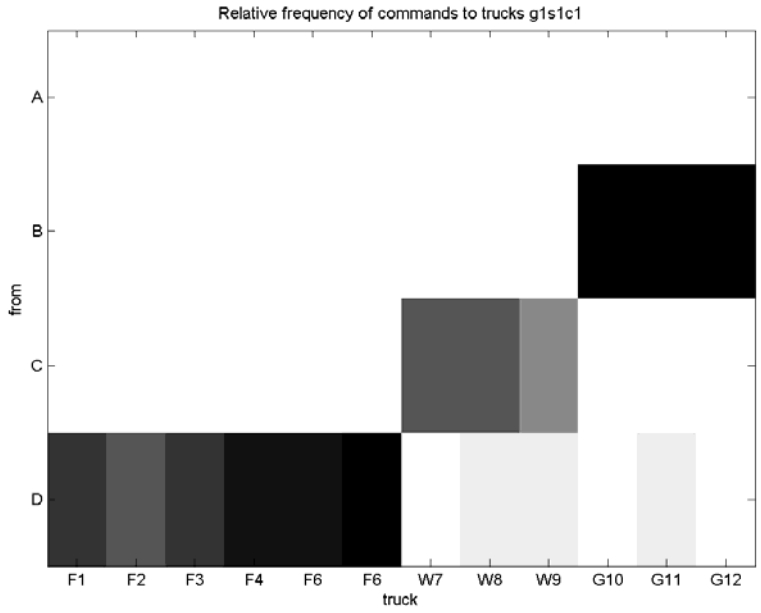
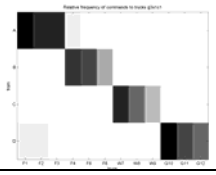
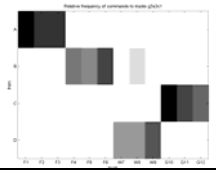
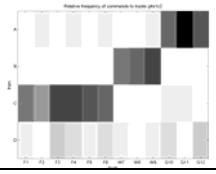
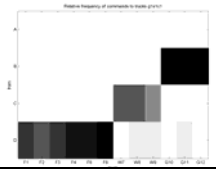
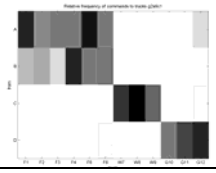
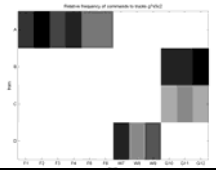
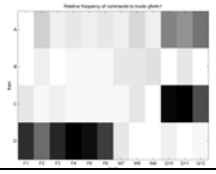
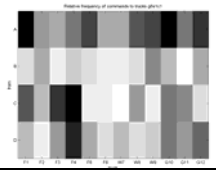


Figure 6: Relative frequency of commands from participants on one team to trucks during one trial: Swedish group 1, scenario 1, team 1.

One of the authors printed all matrices without any identifiers regarding the nationality of the team. The other author clustered them based on patterns in the distribution of the grey tones. We then devised categories that identified each cluster. We found eight different patterns of command structure and truck distribution. Table 3 illustrates and describes the eight different categories of truck distribution found when analyzing all C3Fire trials from all three national groups.

Table 3: Categories of truck distribution in C3Fire

Category – name	Distribution of trucks	Examples
Partitioned according to 'convenience'	The participants command three trucks each. The partition is based on participant name and truck number: Participant A commands trucks 1-3; participant B commands trucks 4-6; participant C commands trucks 7-9, and participant D commands trucks 10-12.	
Partitioned according to 'preference'	The participants command three trucks each. The partition is based on the participants' preferences.	
Assistant	One participant coordinates the other participants' actions through email communication and actively commands the trucks when he finds appropriate. The other three participants divide the trucks according to type.	
Coordinator	One participant coordinates the other participants' actions through email communication. The other three participants divide the trucks evenly according to truck type. The leader interferes occasionally but does not send commands to more than 3 trucks.	
Shared fire trucks	Two participants command the fire trucks together. The third participant commands the gas trucks and the fourth commands the water trucks.	
Shared gas trucks	One participant commands all six fire trucks. Another participant commands the water trucks and the other two participants command the gas trucks together.	
Vague structure	The participants command several trucks but put emphasis on some trucks. In the example we can see that all participants command almost all trucks but that they put emphasis on particular types of trucks. A vague structure can be seen.	
No visible structure	There is no visible structure. Most participants send commands to a large number of trucks.	

There were differences in preferences of truck distribution across the national groups. Table 4 shows the frequency of each category in each national group.

Table 4: Truck distribution in the Swedish, Indian and Iranian groups when engaged in C3Fire

Organization Categories (total sum in parentheses)	Swedish participants	*Frequency %*	Iranian participants	*Frequency %*	Indian participants	*Frequency %*
Partitioned convenience(=33)	33	*0,52*	-	-	-	-
Partitioned preference (=14)	7	*0,11*	-	-	7	*0,22*
Shared fire trucks (=4)	4	*0,06*	-	-	-	-
Shared fuel (=5)	-	-	-	-	5	*0,16*
Assistant (=14)	7	*0,11*	6	*0,38*	1	*0,03*
Coordinator (=7)	2	*0,03*	4	*0,25*	1	*0,03*
Vague structure (=21)	9	*0,14*	4	*0,25*	8	*0,25*
No visible structure (=12)	1	*0,02*	2	*0,13*	10	*0,31*

The partitioned distribution of trucks based on convenience was by far the most frequent type of truck distribution in the Swedish group. The Swedish group also used the partitioned distribution based on preference a number of times, and so did the Indian group to a large extent. In contrast, this type of distribution was never used by the Iranian group. The two types of partitioned distributions represent a well-structured, egalitarian approach to the allocation of responsibilities across participants. The critical observation here is that our Iranian participants did not use this type of distribution, even though they were raised in Sweden. The Swedish propensity to egalitarianism (Hofstede, 1980; Schwartz, 1992, 1994) does not appear to have been instilled to the same degree in these young men who identify themselves as Iranian Swedes. This encourages us to continue look for culturally driven differences in the immigrant population in Sweden.

The Coordinator distribution represents a highly structured and egalitarian distribution of task allocation in which there is a nominal leader. The leader monitors the game and sends emails to the team members about what needs to be done. He sometimes controls the trucks, but no more than three trucks during the whole game. It was used by the Iranians several times and the Swedes during two games. This distribution can be seen as an egalitarian but more controlled structure than the partitioned distribution. In the Assistant distribution, on the other hand, the leader actively commands both participants and trucks. This represents a less egalitarian and more formally hierarchic allocation of responsibilities since the leader often overrides the other participants' commands. It was used by all three groups. Note that the Iranian favored these two types of distribution. The Iranians appear to prefer to have a leader supervising or controlling the other participants' actions.

The two Shared categories represent truly cooperative approaches to the task. There is no clear leader, no coordinator, and no assistant who directs the other participants. The shared fire truck approach was used a couple of times by the Swedes but was not a dominant mode of working for the group. The shared fuel approach was used by the Indians.

The two last truck distributions, vague of no structure, are dominant in the Indian groups. Everyone seemed to drive more trucks than they were allotted. Everyone drove a little bit of everything. It is not clear from these data whether the Indians distrust organization or are truly cooperative or are comfortable with spontaneous chaos.

In sum, the Swedes preferred a non-partial partitioned division of tasks based on convenience rather than preference. On a few occasions there was a leader in the Swedish teams. The Iranians preferred a structure with a leader or coordinator keeping track of the game and the team members' activities. The Indians preferred to share the allocations of tasks and had seldom a team leader.

The experiments using C3Fire indicate that there are differences in both value preferences and task allocation among the three groups. As our experiments continue and our analysis is extended to include

the communication data, we will be able refine these results. If these differences were to be seen in an OSOCC, the implications would likely be dramatic. The contrasts in preferred operation procedures could easily give rise to conflicts in an OSOCC or other command and control setting.

Study 2 – Observation

As a complement to the experimental research, we conducted an observational study of an exercise, run by SRSA (Swedish Rescue Services Agency) and DEMA (Danish Emergency Management Agency). The exercise was designed to train future OSOCC professionals' skills in coordination work and to instill an understanding of the significance and content of OSOCC work in the participants. Two parts of the exercise were observed through exploratory observation: (1) the exercise management, and (2) a team of eight decision makers participating in the exercise. The aim of the observation was to understand how people participating in OSOCC work are trained to stand prepared for an international relief effort.

The exercise was a real-time full-scale exercise simulating the first crucial hours of an OSOCC mission. It lasted for about 24 hours and was part of a seven-day course in operational management. The exercise was conducted on the fifth and sixth day of the course and was considered the grand finale of the course during which the participants were to implement what they had learned during the course. Twenty-four participants from all over Europe participated in the course. During the first day of the course, the participants were divided into three teams. Each team was given a team meeting room. Every day of the course the teams got some information about the exercise and were given the opportunity to prepare for the exercise. When the fifth day and its exercise came, the team had prepared a travel route, technical equipment, task assignments, plan of an sub-OSOCC, contact lists, etc..

The exercise scenario

The scenario postulated that a hurricane had left the infrastructure of a fictional country devastated and its population in despair. The teams' mission was to reach the affected country (Denmark), from the neighboring country (Sweden), and set up a sub-OSOCC. Each team carried out the exercise separately. The three teams did not have contact with one another during the exercise.

The teams were assigned region within the DEMA exercise facility where they were to set up their sub-OSOCCs. The sub-OSOCC's objective was to coordinate the international relief efforts in the allotted region, following instructions from the Main OSOCC (impersonated by the exercise management staff) and other stakeholders (e.g., LEMA, adhering to the OSOCC Concept structure).

The exercise management

The course participants were divided into three teams. For every team, an exercise management Team Manager and an observer was appointed from the SRSA/DEMA staff (see Figure 7). The Team Managers' tasks were to direct the exercise for his/her team. The managers ran the exercise with the help of pre-defined scripted stage directions, consisting of scheduled events, such as phone calls, e-mails, actions and incidents to which the manager had to expose the team. The exercise was a truly dynamic process, where the predefined actions to be played out in the exercise functioned only as guidelines for the exercise. Some events were carried out according to the script and other events were played out when appropriate as determined by the actions that had been taken by the teams. The team managers therefore continuously assessed the situation and the team's performance. If more or less action was needed, the exercise was modified accordingly. The dynamic simulated situation was illustrated on maps and schedules that were available for all personnel on the exercise management team.

Most events, such as phone calls and emails, were implemented by the manager himself. Some events and actions were implemented with the help of actors, ensuring that the exercise would be realistic for the exercise participants. The actors were mainly SRV/DEMA professionals of various backgrounds who had had no previous contact with the exercise participants. The actors impersonated important stakeholders that the OSOCC personnel are likely to run into during a mission, such as police and

military officers, customs personnel, UNDAC and LEMA people, media representatives, international relief team managers, and more.

Each team had an appointed observer. The observers had many years experience in international relief operations and OSOCC work. They functioned as evaluators of the teams' performance, as instructors, and as assistants to the teams. Each observer followed his team during the whole exercise. Several times during the exercise the observer did a "time out" in the exercise and commented on the team's performance up to that point. He pointed out actions that were successful and less successful and encouraged the team to reflect on their performance, team dynamics, and future actions. The observers continuously reported to the Team Manager and thereby helped the Team Manager to assess the current situation.

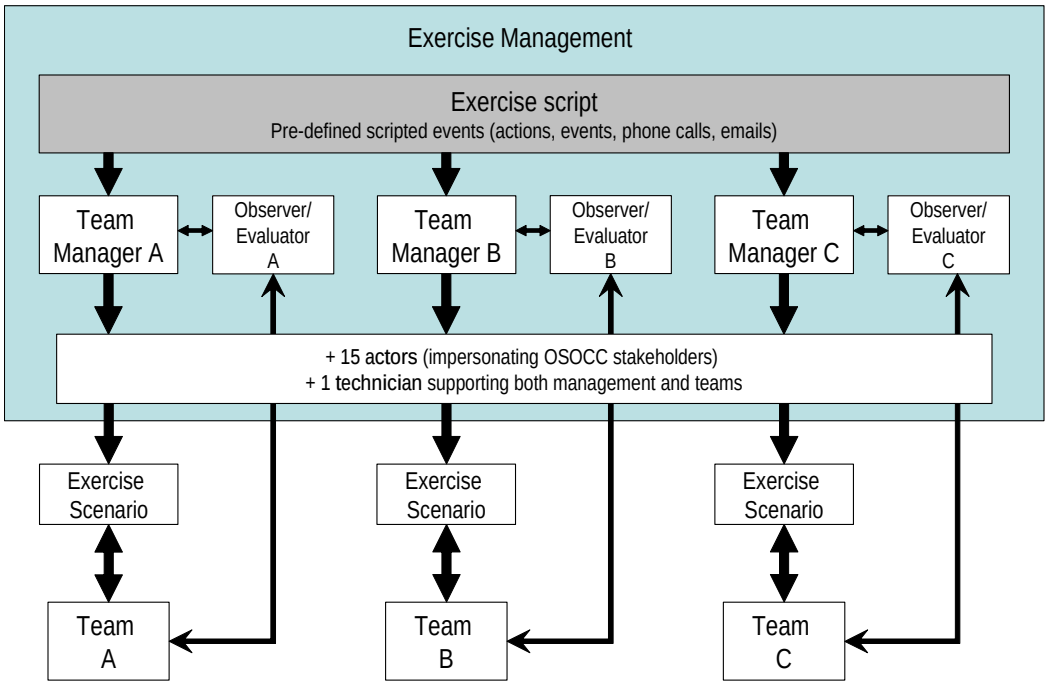


Figure 7: The exercise structure. Each team's exercise events were directed by a Team Manager, and each team was followed by an observer/evaluator. The events were played out according to an exercise script and in relation to teams' responses to earlier events. The Team Manager and/or actors played the events, resulting in a realistic and dynamic scenario. A technician supported the exercise management and the teams' handling of technical equipment.

The management staff also included a technician, specialized in all kinds of technical equipment needed during a mission, who functioned as technical support for both the management and the participating teams.

The exercise and course as a whole concluded with an evaluation session of the exercise with all three teams together. Since all teams had experienced the exercise differently, with some events left out or added in the individual teams, this session functioned mainly as a group discussion in which overall performance of the teams was discussed, rather than specific team members or episodes. During this session, the team members were encouraged to share their experiences during the exercise.

The observed team

One of the authors followed one of the teams as their "shadow", i.e. sat in during their meetings two days before the exercise and inactively took part of the exercise. The team was told that the shadow's intention was to learn more about OSOCC work.

There were eight team members from eight different countries. These people had not only different national backgrounds (all Europeans), but also different professional backgrounds (military, fire brigade, civilian), and different experiences of similar missions.

The exercise started at Revinge in Sweden where the team received two cars, some office supplies and money. The team was also given technical equipment in the shape of laptop computers, satellite phones, mobile phones, a printer/scanner, and more. During the days before the exercise, the team was given maps, some information about the hurricane, and background information about the affected region. From the location in Sweden, the team had to drive to a facility in Denmark, an area where they had never been before. As a part of the exercise, the team got stuck in customs, which was the first trial of the exercise. After solving the customs problem, the team contacted the Main OSOCC (the exercise management) and asked where to set up their sub-OSOCC. The team was given instructions to set up their sub-OSOCC at a specific location (in DEMA exercise facilities). When arriving to the exercise facility, the team was allotted a locale for their sub-OSOCC and some furniture.

From their sub-OSOCC, the team's tasks were to assess the situation in a region of the affected country allotted by the Main OSOCC, to contact important stakeholders and to help coordinate the international relief operations.

Results

The exercise management

Observing the exercise management gave several opportunities for informal interaction with SRSA/DEMA personnel that have been active during OSOCC or similar work. Personal communication with people who had been part of international relief efforts revealed that it is crucial to know the local norms for behavior if one wants to be successful. To show up too early or late to a meeting can be devastating for the OSOCC's reputation. What is considered too late/early is culturally specific. Leadership styles were also reported to differ and pose barriers to efficient cooperation. For example, one Swedish fire chief, who usually works as a team leader when on international missions, reported that he usually tells his Swedish personnel not to be informal with him when others can see it. The Swedish leadership style is known to be egalitarian and "friend-like". He had noticed that when his Swedish employees talked to him the way they do on Swedish ground, people from other cultures sometimes considered him to be without authority and lost their respect for him. This is an example of Power Distance (Hofstede, 1980). Swedish employees do not expect much difference in power between themselves and their leader. Many decisions are based on consensus in the group and are often compromises between differing views. In many other cultures, however, a large difference in power is expected between the person in charge and the followers. In these cultures, the leader often has the right to decide without taking his employees' thoughts into account.

The observed team

The observation suggests that cultural differences are salient and acknowledged by the participants. There seemed to be clear differences in leadership and communication styles that would be interesting to study further. As an example, the team was led by a fire chief from northern Europe, here called 'N'. He seemed comfortable as the leader and directed the team meetings with the help of highly organized lists of topics to be discussed. N's careful planning and egalitarian leadership are exemplars of moderate uncertainty avoidance (Hofstede, 1980) and of a preference for security (Schwartz, 1992, 1994).

During one meeting, the group was asked by the leader to prioritize in what order the topics where to be brought up for discussion. This took time from the meeting but ensured that everyone knew what was to be discussed and in what order. After discussing the first two bullets on their list, a man from the south of Europe, here called 'S', who also is a fire chief, interrupted the discussion. S wanted to move on to the fourth topic on the list, a topic that was related to his tasks. N wanted to keep to the pre-defined list of topics, but S insisted on discussing his topic with the motivation that "It will be quick". A discussion started about whether or not the group should continue to discuss the current topic or move

on to S's topic, resulting in a change of topic. S's unwillingness to follow the pre-defined agenda and moving directly to his own topic can easily be interpreted as an instance of strong uncertainty avoidance and an indicator of a preference for achievement. According to Hofstede (2005), people from weak uncertainty avoidance cultures are often perceived by others as quiet and controlled, whereas people from strong uncertainty avoidance cultures can be perceived as busy, emotional and aggressive. These tendencies were illustrated by the two fire chiefs' behavior.

Several members of the team had severe difficulties speaking and understanding English, which was the official language during the course and exercise. Both during the course and especially during the exercise, confusion arose during meetings due to language misunderstandings. This is likely to generalize to real international command and control situations, such as OSOCC missions. It is not possible to tell in advance which people will be part of the OSOCC, where they come from and what background they have. It is therefore not certain that they have a shared language. This creates problems. If the persons in coordination or command positions cannot understand each other, how can they ensure effectiveness and efficiency?

Due to the differences in professional backgrounds (military, fire brigade, civil authority), the team members seemed to have different expectations for meeting behavior and the exercise itself. The exercise was structured to provide the participants with the opportunity to implement procedures learnt during the operational management course. However, familiar procedures imported from their normal work seemed to surface and rule their work. As a result, on several occasions, some participants complained about and appeared to have conflicts with the team's work procedures. When team members from different cultures import differing procedures, conflicts are bound to arise. When combined with differing leadership styles and planning behavior, adhering to idiosyncratic procedures can lead to serious misunderstandings. This was reflected in the team's meeting behavior. Differences in time frame for planning and norms for behavior and communication resulted in ill-structured meetings with much confusion about tasks and procedures. These types of conflicts are likely to occur during real OSOCC missions and during command and control missions in general. They underscore the importance of shared training and the need to define clusters of cultural differences.

Initial Clusters of Culturally-Driven Differences in Expectations and Preferences for Collaborative Decision Making

We interpret our data to reveal that the three cultures diverge along four dimensions: (a) the presence or absence of a clear team leader, (b) their attitude toward ambiguity (or more generally, risk), (c) their expectations for and tolerance of conflict in the course of task performance, and (d) their methods for allocating roles and tasks across team members. Leadership style directly influences the propensity for hierarchical organization. Risk aversion (as revealed by the team's attitude toward ambiguous information) focuses that organization's efforts. Leadership style and risk aversion resonate with Hofstede's (1980) power distance and uncertainty avoidance. The third dimension, expectations for and tolerance of conflict, refers to attitudes and behaviors directed at establishing and maintaining both intra-team and inter-team harmony. The last dimension, role allocation, is novel and addresses the process and rigidity of team formation.

The first three dimensions are essentially bimodal, e.g., a cultural group either prefers to have a team leader or does not. Conflict aversion and ambiguity aversion appear to be closely linked in our subject pools. Cultural groups that seek to clarify ambiguous information also seek to minimize discord. Other groups appear to perceive ambiguity and discord as normal parts of team processes.

The bimodality of the first three dimensions and the apparent alignment of attitudes toward ambiguity and conflict suggest a natural partitioning of our results into four clusters that can be summarized in the 2 x 2 matrix shown in Table 5. The clusters in the upper row represent teams with clear leaders. Teams without leaders form the clusters in the bottom row. The clusters in the left-hand column represent teams that actively seek to promote harmony and to clarify uncertain information. Teams that are

relatively ambivalent to conflict and ambiguity occupy the clusters on the right. The multiple methods for role allocation are distributed across the clusters independently of the matrix structure.

Table 5: Clusters of patterns of behavior associated with cultural differences in collaborative decision making

Risk-Averse Egalitarian Hierarchies • The team has a leader who directs/coordinates task performance • Roles and tasks are clearly partitioned and largely adhered to • The leader strives to avoid conflicts and eliminate ambiguity	**Risk-Tolerant Hierarchies** • The team has a leader who serves as an assistant • Roles and tasks flex depending upon task demand • Conflicts and ambiguity are perceived as natural and inevitable parts of distributed work
Risk-Averse Egalitarian Teams • No clear leader • Roles and tasks flex depending upon task demand • The team jointly works to avoid conflict and ambiguity	**Risk-Tolerant Inchoate Groups** • No clear leader • Ambiguous roles • Multitasking (no clear distribution of tasks) • Conflicts and ambiguity perceived as natural and inevitable parts of distributed work

Table 5 is an initial attempt to formulate a clustering of culturally driven differences that can be expected to influence collaboration in international command and control settings. We present these clusters with several disclaimers. First, it is likely that the clustering shown here will evolve as our work progresses. Second, it is important to remember that our results are strictly valid only for the groups we studied; they will not generalize to everyone from the nations of Sweden, Iran, and India. Nevertheless, our data suggest clusters of differences that may pose barriers to efficient and effective cooperation in international teams.

The cluster in upper left of Table 5 reflects preferences for a relatively traditional hierarchic team structure. The team has a leader who oversees all of the team's activities. Each team member has a clear role that is determined either by convenience or preference. That partitioning of roles is largely adhered to. Part of the leader's job is to promote internal harmony and minimize uncertainty. This cluster of attitudes is revealed by the 'coordinator' type of command matrix during the C3Fire experiment (see Tables 3 and 4). The clearest instance of this cluster was witnessed during the observational study. Fire chief N's careful planning in cooperation with his team, during which he tried to involve all participants of the team in discussion, illustrates an egalitarian leadership style. As an aside, it is important to note that none of the clusters represents an archetypically top-down 'militaristic' hierarchy. Even the most traditional clustering displays a refreshing egalitarianism.

The cluster in the upper right reflects preferences for a somewhat hierarchic team. The team has a leader but roles, while specified, are not rigid. Team members can take on or drop responsibilities in response to changes in the task demands. The option to flex may be associated with a willingness to tolerate both ambiguous information and inter-team conflict. Examples of flexible behavior are seen in the 'assistant' type of command matrix (see Tables 3 and 4). Sending commands to trucks that had been assigned to another participant has two implications. First, it reveals a willingness to let roles flex. Second, it raises the possibility of conflict (e.g., "Don't drive my trucks!") and implies that such conflicts are an acceptable part of team functioning.

The cluster in the lower right reflects preferences for behaviors that appear to some to be chaotic. As reflected by the 'vague structure' and 'no visible structure' types of command matrix (see Tables 3 and 4), these teams have no clear structure and tolerate considerable ambiguity. Harmony is not a concern.

It is important to note that many of these inchoate groups suppressed the fires quickly. Thus there appears to be no correlation between the reification of team of structure and degrees of success.

The cluster in the lower left reflects preferences for a leaderless organization of equals. As reflected by the 'partitioned' and 'shared' types of command matrix (see Tables 3 and 4), roles are determined either by convenience or preference but are allowed to flex in response to the dynamics of the situation. The team works together to promote internal harmony and minimize uncertainty. This cluster appears to be strongly influenced by the Scandinavian model of social cooperation, mutual respect, and harmony.

Finally, and somewhat reluctantly, we identify our national groups with these initial clusters. The majority of the northern Europeans observed during the exercise appeared to be comfortable working in a Risk-Averse Egalitarian Hierarchy. We suspect that Fire chief S, from southern Europe, might have been more comfortable with a more formal and less egalitarian hierarchy.

Our Iranian participants, who were raised in Sweden by parents who fled the ouster of the Shah, appeared to favor a Risk-Tolerant Hierarchy. In contrast, our Swedish participants with their many generations of Swedish ancestors gravitated to Risk-Averse Egalitarian teams. The Swedes were alone in preferring a partitioned structure. While the Iranian group's score on the time horizon questionnaire suggests that they plan much as Swedes, their score on the conflict avoidance questionnaire suggests they do not shun conflict as much as Swedes. The Iranians prefer hierarchies; the Swedes do not. The contrast between these two sets of young adults raised in Sweden is somewhat remarkable and can be attributed to their diverse cultural heritage.

Finally, our Indian participants invariably developed relatively ill-formed groups with no clear leaders and no partitioning of roles. This result obtained even when all members of the group came from the same state (i.e., Madhya Pradesh or Tamil Nadu) and spoke the same language (i.e., Tamil or Urdu). The Indian groups' performance during the C3Fire sessions revealed little taste for hierarchical structure. Their scores on the time horizon questionnaire suggest that they did not concern themselves much with long-term planning, but rather were caught up in the moment. Our data suggest that the buzzing confusion of the Indian marketplace reappears in collaborative groups.

It is not difficult to imagine how these differences could compromise cooperation in OSOCC or other command and control centers if individuals who naturally align themselves with different clusters were to work together. If these and other cultural influences can be pinpointed, clustered, and illustrated for OSOCC staff, the OSOCC might be better prepared to deal with conflicts based on cultural differences. We are continuing our research and will continue to refine our definition of culturally-driven clusters of approaches to collaborative decision making.

References

Artman, H. & Wearn, Y. (1999). Distributed Cognition in an Emergency Co-ordination Center. *Cognition, Technology & Work, 1(4),* pp. 237-246.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior, 9*(2-3), 171.

Brehmer, B. (2005) Micro-worlds and the circular relation between people and their environment, *Theoretical Issues in Ergonomics Science*, 6, 1, 73-93.

Chu, P.C., Spires, E. E., Farn C. K. and Sueyoshi, T. (2005) Decision Processes and use of decision aids – comparing two closely related nations in East Asia. *Journal of Cross-Cultural Psychology, 36(3),* pp. 304-320.

Chu, P.C., Spires, E. E., and Sueyoshi, T. (1999). Cross-Cultural Differences in Choice Behavior and Use of Decision Aids: A Comparison of Japan and the United States. *Organizational Behavior and Human Decision Processes, 77(2),* pp. 147-170.

Costa, P.,T., Jr and McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor (NEO-FFI) Inventory professional manual.* Odessa, FL: PAR Inc.

Daun, Å. (1999). It's Rude to Interrupt! Swedish Communication Styles. In Daun, Å. & Jansson, S. (Eds) *Europeans: Essays on Culture and Identity*. Nordic Academic Press, Lund, Sweden.

Daun, Å. (1998). *Svensk mentalitet.* Rabén Prisma, Stockholm, Sweden.

Di Luzio, A.D., Günthner, S. & Orletti, F. (Eds) (2001). *Culture in Communication -Analyses of intercultural situations.* John Benjamins Publishing, Amsterdam.

Duranti, A. (1997). *Linguistic anthropology.* Cambridge University Press.

Dörner, D. & Schaub, H. (1994). Errors in Planning and Decision Making and the Nature of the Human Information Processing. *Applied Psychology: An international review, Special Issue on Human Error,* pp. 433-453.

Granlund, R. (2002). *Monitoring Distributed Teamwork Training.* Unpublished PhD Thesis, Linköpings Universitet, Linköping, Sweden.

Granlund, R. (2003). Monitoring experiences from command and control research with the C3Fire microworld. *Cognition, Technology & Work, 5*(3), 183-190.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly, 2*, 205-227.

Hofstede, G. (1980). *Culture's Consequences – International differences in work-related values*. Thousand Oaks, CA: Sage Publications

Hofstede, G. (2001). *Culture's consequences: Second Edition.* Thousand Oaks, CA: Sage Publications.

Hofstede, G & Hofstede G. J. (2005). *Cultures and organizations: software of the mind.* 2nd Edition.McGraw-Hill, London.

Johansson, B., Persson, M., Granlund, R., & Mattsson, P. (2003). C3Fire in command and control research. *Cognition, Technology & Work, 5*(3), 191-196.

Johansson, B. (2005). *Joint Control in Dynamic Situations.* Unpublished PhD Thesis, Linköpings Universitet, Linköping, Sweden.

Kim, H. & Markus, H.R. (1999). Deviance or Uniqueness, Harmony or Conformity? A Cultural Analysis. *Journal of Personality and Social Psychology, 77(4),* pp. 785-800.

Mann, L., Radford, M., Burnett, P., Ford, S., Bond, M., Leung, K., Nakamura, H., Vaughan, G. & Yang, K. (1998). Cross-cultural Differences in Self-reported Decision-making Style and Confidence. *International Journal of Psychology*, 33(5), 325-335.

OCHA's official homepage: http://ochaonline.un.org

Schwartz, S. H. (1992). Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. *Advances in Experimental Social Psychology, 25,* pp. 1-65.

Schwartz, S. H. (1994). Are There Universal Aspects in the Structure and Contents of Human Values? *Journal of Social Issues, 40 (4),* pp. 19-45.

Schwartz, S. H. & Bilsky, W. (1987). Toward a psychological structure of human values. *Journal of Personality and Social Psychology, 53,* pp.550-562.

Smith, P.B. & Bond, M. H. (1999). *Social Psychology Across Cultures.* London, Prentice Hall Europe.

Triandis, H. C. (1996). The Psychological Measurement of Cultural Syndromes. *American Psychologist, 51(4),* pp.407-415.

United Nations (2000). United Nations Disaster Assessment and Coordination UNDAC Field Handbook. Third Edition. Available 2006-06-12 at http://www.reliefweb.int/rw/lib.nsf/db900SID/LGEL-5EGP65/$FILE/undac-manual-2000.pdf?OpenElement

United Nations (2002). OCHA Orientation Handbook. Available 2006-06-12 at http://www.reliefweb.int/rw/lib.nsf/db900SID/LGEL-5R8DWY/$FILE/ocha-orient-2002.pdf?OpenElement

Woltjer, R. (2005). *On How Contraints Shape action.* Unpublished Licentiate Thesis, Linköpings Universitet, Linköping, Sweden.

Running head: ECONOMICS AND EMOTION

Economics and Emotion: Institutions Matter

Kip Smith

Linköping Institute of Technology

John Dickhaut

University of Minnesota

Address correspondence to Kip Smith, Industriell arbetsvetenskap, Institutionen för Konstrucktion och produktionsteknik, Linköpings universitetet, 581 83 Linköping, Sweden; e-mail: kipsm@ikp.liu.se. Phone: (011) 46 13 282764 Fax: (011) 46 13 282579

Accepted for publication in Games and Economic Behavior.

The paper has benefited from presentations before the Economic Science Association, the American Economic Association, the Society for Judgment and Decision Making, the Neuroeconomics Conference at the University of Minnesota, the University of Minnesota Accounting Department, and from comments by Aldo Rustichini, Douglas Bernheim, Antonio Rangel, Colin Camerer, Paul Slovic, John Ledyard, Jack Kareken, and Vesna Prakister. This work was supported in part by grants from the National Science Foundation and the Department of Defense Multidisciplinary University Research Initiative (MURI OPUS).

Abstract

In two different types of institutions, English- and Dutch auctions, we collect heart rate data, a proxy for emotion, to test hypotheses based on findings in neural science about the effect of emotion on economic behavior. We first demonstrate that recording heart rates does not distort prices in these auctions. Next we consider whether knowledge of the intensity of a participant's emotional state improves our ability to predict price setting behavior beyond predictions of price based on usual economic variables. Our answer is that "Institutions Matter." In the Dutch (English) auctions we find (no) evidence that knowledge of emotional intensity affects our ability to predict price setting behavior. We then entertain the proposition that the cardiac system is an information system that processes economic events. We are able to show that this hypothesis is consistent with our observations and furthermore that the processes differ across institutions.

Economics and Emotion: Institutions Matter

The purpose of this paper is to extend what has been a core practice of experimental economics, namely asking if the truth of an economic proposition depends on the market institution in which it is studied. In particular, we ask if the impact of emotion is the same when similar economic choices are compared across institutions. The choices we examine are the choices of prices setters in English and Dutch auctions (Vickrey, 1961). We ask the degree to which knowledge of the intensity of the price setter's emotional state will give us additional insight into these choices after controlling for the usual economic variables assumed to affect them. We find the role of emotion depends on the form of market institution studied.

Early in the development of experimental economics, market institutions were found to matter. Plott (1982) demonstrated that monopoly pricing depends on the institution in which trade occurs. A double auction with a single seller is more competitive than a posted offer auction with a single seller. Kagel (1990) showed that individual behavior can differ across institutions with identical equilibrium allocation implications, namely the English and Second-price auctions. Smith (1982) has outlined how the endeavor of laboratory experiments can be viewed as the empirical study of a mechanism's ability to implement allocation goals such as surplus maximization.

Two recent studies relevant to this paper show that assessments of individual preferences can in fact change across institutions even when these assessment are made within minutes of one another. Using standard methods of assessing preferences in the English, First-price, and the Becker, Degroot and Marschak mechanisms (BDM), Berg, Dickhaut and McCabe (in review) demonstrate that individual preferences are more likely to be risk preferring in the English auction, risk averse in the First Price auction, and somewhat in between in BDM. In a related paper, Isaac and Jackson (1994) vary the BDM and First-price auction and find that assessed preferences in these two institutions also change. These results lead to the concern that either preferences may depend on the institution in which they are studied or something else is at work that is not captured by usual representations of preference.

Support for these concerns can be found in recent neuroscience research. Antonio Damasio and his colleagues have argued that emotions may be playing a fundamental role in choice (Bechara, Damasio, Damasio, & Anderson, 1994; Bechara, Damasio, Tranel, & Damasio, 1997; Damasio, 1994, 1996, 1999). They contrast results from experimental participants with no history of brain damage or psychological disease ("normals") with results from patients with damage to the ventromedial prefrontal portion of the brain[1]. The Damasio team found that damage in the ventromedial area of the brain made patients less susceptible to the influence of negative stimuli such as losses associated with draws from decks of cards. Their work implicated the role of emotions by examining heart rate and galvanic skin response, two psychophysiologic proxies for emotion. Heart rate, the proxy for emotions that we use in this study, has been frequently studied for its ability to reflect emotional intensity (e.g., Ekman, Levenson, & Friesen, 1983; Lane & Schwartz, 1990; Lo & Repin, 2002; Sinha, Lovallo, & Parsons, 1992).

In this study we move the study of emotion in choice to real economic institutions, Dutch and English auctions. In particular we focus on the choices made in these auctions that determine prices. We apply the usual economic models used in the effort to predict pricing behavior in these auctions. Further, we ask if a limited knowledge of the relative intensity of emotion, as provided by heart rate, augments our ability to predict price setting behavior.

We begin this paper with a review of the two institutions, the English and Dutch auctions, and with an account of how the processing of economic and emotional information may modulate heart rate. These reviews motivate hypotheses on the differential role of emotion on price formation across these trading institutions. In the methods section, we discuss data collection, dependent measures, and data reduction.

---

[1] The ventromedial prefrontal cortex has independently been shown to be active during the process of choice (Dickhaut, McCabe, Nagode, Smith, & Pardo, 2003; Glimcher, 2002; Rustichini, et al., this volume; Smith, Dickhaut, McCabe, & Pardo, 2002).

The presentation of results has four parts. We first show that the presence of heart rate monitoring equipment does not distort the typical findings – group and individual behavior - in these auctions. Then we test whether we can improve on the standard economic models and make better predictions of prices in these auctions by augmenting them with the measurement of heart rate just prior to the determination of price. Our interpretation of the results of these tests are tempered by our understanding of the auction in which the choices occur. Emotions play a role for a significant subset of participants in the Dutch auction but not in the English auction. Finally, we turn to the cardiac system as an information processing system of economic signals. We find evidence for such a system and that the nature of processing varies across institutions.

Differential emotion across trading institutions

In this section we describe how English and Dutch auctions were implemented in our study and motivate why emotions (as proxied by heart rate) can differentially affect pricing behavior in these auctions.

Auction structure

As implemented in our laboratory, the auctions are represented by a circular clock on each of the four participants' computer screens. In the English auction, the clock begins at $0 (in the noon position) and increases continually to a maximum of $2.50 (again at noon) at a rate of $0.05/second. Each participant starts out <u>in</u> the auction and presses a computer key to drop out of the auction. The last participant in the auction wins the auction at the price at which the next to last (third) participant exits. The last person to exit is the price setter and that participant's price is the <u>exit price</u>. In contrast, the clock in the Dutch auction begins at a price higher than anyone is willing to pay for the item ($2.50 in our auctions). Thus each player is initially out of the auction. As the price decreases the participant decides whether to enter the auction by pressing a computer key. The first participant to press a key wins the auction. The first person to enter is the price setter and that participant's price is the <u>entry price</u>. Across institutions, to entice the participants to want to apply any profit motives they have, the computer generates a random draw of a dollar and cents amount, called the participant's <u>value</u>, from $0.00 to $2.50, prior to the start of each auction. All amounts are equally likely in each draw. There is a separate and private draw for each participant. Should a participant win the auction, the

difference between the participant's private value and the price in the auction is the participant's profit in that auction. For these reasons the two auctions are best described as private value English and Dutch clock auctions.

In the English auction reasoning is straightforward: a participant who wins an auction will make a profit as long as all the other participants exit at a price less than the winner's value; the winner loses money when the last exit occurs at a price greater than the winner's value. Therefore, we expect participants to stay in the auction until the price reaches their values and then exit. Further, they do not have to consider what other participants are doing to decide when to exit.

On the other hand, participants in the Dutch auction knows that it is possible to achieve a positive profit only after the clock has reached their values and that the profit will increase the longer they wait to enter. However, the participants cannot be sure of the profit while they wait because by staying out they are vulnerable to other participants' winning the auction. The decision regarding when to enter may be affected by fears of what the others might do. A pair of emotional factors may influence pricing behavior here. First, the realization that a profit can be realized may produce a euphoria. Second, the problem of not knowing what others might do may induce anxiety. The euphoric effect may lead to entering the auction later, the anxiety to entering earlier.

Emotions and heart rate

We now turn to a more complete scientific explanation underlying why we expect to see differences in emotional behavior across auction institutions and why we expect to be able to detect this differential by monitoring participants' heart rates. There are two main routes from the brain to the heart that enable visually-acquired economic information (e.g., auction data on computer screens) to have an essentially instantaneous impact on the rate at which the heart beats. The upper network projects dorsally (across the top of the brain and head) from visual areas in the occipital cortex to areas in parietal and frontal cortex where, we assume, value and current prices are represented and become the focus of conscious cognition. The lower network is ventral (across the bottom of the brain). It projects to the ventromedial prefrontal cortex, the region at the base of the brain that lies behind the eyes. The ventromedial prefrontal cortex has

been shown to be critically involved in the processing of visceral information and emotion both with and without the mediation of consciousness (e.g., Bechara et al., 1994). Both networks communicate with the hypothalamus, the small but pivotal subcortical region that controls the two divisions of the autonomic system, the sympathetic and the parasympathetic.

The interplay of sympathetic and parasympathetic dominance and withdrawal controls the rate at which the heart beats. The feedback system linking the ventromedial prefrontal cortex and the hypothalamus (e.g., Bechara, Damasio, Damasio, & Lee, 1999; Dolan, Fletcher, Morris, Kapur, Deakin, & Frith, 1996) allows the representation and processing of emotional information to interact with the rate at which the heart beats. The dorsal (upper) network is known to support a variety of information processes including, but not limited to, calculation (e.g., Menon, Rivera, White, Glover, & Reiss, 2000), working memory (e.g., Marshuetz, Smith, Jonides, DeGutis, & Chenevert, 2000), and decision making (e.g., Bechara, Tranel, Damasio, & Damasio, 1996; McCabe, Houser, Ryan, Smith, & Trouard, 2001; Rogers et al., 1999). We assume that the dorsal stream of processing drives most auction behavior. Some of this behavior is likely to be relatively effortful (e.g., price setting choices in the Dutch) while some is likely to be relatively effortless (e.g., the English). When cognitive processes require relatively high levels of effort, the sympathetic division of the autonomic system becomes dominant and causes the heart to beat faster (Aasman, Mulder, & Mulder, 1987; Backs, 1995). In contrast, when processing is relatively easy, the parasympathetic system becomes dominant and causes the heart to slow down.

The ventromedial prefrontal cortex is likely to receive information about the auction from two sources, from visual cortex via the ventral network and from visceral information about the status of the body's response to the auction environment. This interaction of visual and visceral information may or may not give rise to gut feelings, that is, to a conscious representation of an emotional response. Damasio (1994, 1999) argues that the ventromedial prefrontal cortex organizes experience and related emotions such that, when a decision situation is encountered, its characteristics (e.g., values and prices) elicit recall of past emotional experiences consistent with the situation. The reemergence of these "emotion settings," in turn, may influence how the choice is made (e.g., Loewenstein, 1996; Mellers, Schwartz, Ho, & Ritov, 1997, 1999; Slovic, Finucane, Peters, & MacGregor, 2002). Rustichini et. al. (this volume) find evidence that for

some economic decisions choice behavior appears to be shifted more toward upper regions of the brain such as the parietal and frontal lobes, a result which makes plausible the notion that in other economic choices, such as the English and Dutch auctions, decisions may invoke different processes and result in different choice observations.

Heart rate and emotions have been examined in a number of studies (e.g., Collet, et al., 1997; Levenson, Ekman, & Friesen, 1990). Usually the participant is put in a setting that is designed to induce an emotional response, e.g., viewing pictures, reading text, or acting a part in a play. While engaged in the task, the participant's heart rate is monitored and questions are asked to assess the participant's emotional state. Typical results are shown in Figure 1. The horizontal axis shows the seven standard categories of emotion. The vertical axis indicates the relative elevation of heart rate above a baseline measurement. Many of the basic emotions induce a change in heart rate. We measure heart rate in our study using a standard three-electrode electrocardiogram.

The effect of context on brain activation

There is ample precedent for suggesting that minor contextual changes (e.g., different rules for auctions) may trigger profoundly different patterns of neuronal activation (along the dorsal or ventral path) which, in turn, may be reflected in changes in observable behavior (price formation). For example, consider the Stroop task (Stroop, 1935). In the Stroop task, the stimulus is a word that names a color, e.g., the word red. The contextual change is provided by the color of the ink that the word is printed in. The ink may or may not match the color's name. For example, if the word is red it might be printed in red ink or green ink. The participant's task is to say the word that corresponds to the color of the ink. It has long been known that both reaction time and accuracy are strongly influenced by whether or not the ink matches the word (for a review see Milliken, Lupianez, Debner & Abello, 1999). Pardo, Pardo, Janer, and Raichle (1990) and Carter, Mintun, and Cohen (1995) used positron emission tomographic (PET) imaging to document that this minor change in context evokes different patterns of neuronal activation. In particular, Pardo found that the ventromedial prefrontal area is preferentially activated in the no match condition.

Pardo's finding has implications for our study of price formation across auction institutions. The Stroop task preferentially activates the ventromedial cortex, the part of the brain that processes visceral, emotional information, when the experimental stimulus presents a conflict. We propose that the environmental differences between the Dutch and English auction also initiate different neuronal networks and that this differential processing leads to different pricing behavior in these auctions. The proposition here is that the English auction, as a collection of stimuli, induces less emotional processing than the Dutch auction and that this difference can in principle be detected indirectly by measuring heart rate.

Method

Participants

Twenty-four undergraduate students (ages 18 to 29, mean age 22.4; 14 females, 10 males), in groups of four, participated in a block of 60 Dutch auctions followed by a block of 60 English auctions. They received a $5.00 show-up award and their cumulative earnings from 120 auctions.

Procedure

Participants sat at separate computer terminals surrounded by privacy barriers. After reading the instructions and signing the consent form, they received five practice Dutch and five practice English auctions. They then played 60 Dutch followed by 60 English auctions. Upon completing both sets of auctions, participants privately received their cumulative winnings. The mean amount won in the auctions was $12.35 with a high of $18.60 and a low of $7.25.

At the beginning of each auction, the computer terminal privately revealed the participant's value for that auction (i.e., the amount the experimenter will pay the participant for owning the asset at the end of the auction). For each participant, this value was randomly drawn from a uniform distribution, $0.00 to $2.50. The random nature of the endowment and its distribution were public knowledge. Values during the auctions remained private information.

Figure 2 shows the computer display used for both types of auctions. At the start of the auction the outstanding price was $0.00 in the English and $2.50 in the Dutch and was

represented by the noon position of the clock. Throughout the experiment, the current price was shown graphically using the clock in the center of the display and numerically (in cents) in the lower right corner of the screen. In the English (ascending price) auction, shown in Figure 2, the clock swept clockwise at a constant rate, five cents per second. In the Dutch (descending price) auction, the clock swept counterclockwise at five cents per second. The display simultaneously changed both the numerical and graphical representations of price. The bottom left corner of the screen indicated the participant's cumulative profits (in dollars) and current value (in cents), both of which remained private information. In contrast, current prices and the number of participants remaining in the auction were public information. The top right corner of the screen displayed the number of participants remaining in English auctions.

Heart rate

Three-electrode electrocardiographic data were recorded using 3M red dot silver/silver-chloride bio-potential electrodes, shielded leads, and four Biopac Systems ECG100C preamplifiers. A separate preamplifier was used for each participant. The settings for the preamplifiers include a low-pass filter and a 50 dB 60 Hz notch filter. A proprietary software program (Pearson Technical Software, Minneapolis, MN) running on a 300 MHz Wintel PC sampled the ECG signals at 1000 Hz. This program uses a dual threshold technique to identify R-wave spikes and filter out movement artifacts and other sources of noise in the signal. It captures the times of the R-waves spikes (the electrical signal that triggers the contraction of the ventricles) and appends packets of 100 spike-times to separate files for each participant. Upon completion of data acquisition, each file contains an hour-plus-long time-series of the times when the heart beats. The 24 time series of heart-beat times, one per participant, are the raw data for our analysis of heart rate and auction structure.

The analysis consists of a comparison of the mean values of heart rate in four-second intervals on either side of an auction event. The means are calculated in the following manner. First the time series of observed heart-beat times was converted to a time series of instantaneous heart rate (in beats per minute, bpm) so that larger values indicate more rapid rates. Second, a linear regression was fit to each time series individually to identify and remove any long-term (generally decreasing) trend that might be associated with accommodation to the experimental

setting and/or fatigue. The residuals from the regressions were then standardized within each participant to create 24 time series of z-scores. All statistical analyses were performed on the z-scores of residual heart rates. The standardization allows aggregation across participants under assumptions of independence across participants and auctions.

Results

Replication of price behavior in auctions

Figure 3 plots of the difference between the exit price (the price at which the price setter exited) in the English auction and the price setter's value against the sequence number of English auctions. Of the 360 observations, 319 plot within 8 cents of the predicted value. This result aligns with previous findings (e.g., McCabe, Rassenti, and Smith, 1990, 1991).

Figure 4 is a plot of the price of the winning entry in the Dutch auction against the winner's value. The upper line has a slope of 1.00 and represents entry at the participant's value. Such behavior would produce no profit. The lower line has a slope of 0.75 and represents the predicted price assuming risk neutrality in a four-person auction. Of the 360 observations, 327 (91%) lie between the two lines, consistent with the assumption that participants were risk averse. This result also aligns with previous findings.

We infer from the data shown in Figures 3 and 4 that the presence of the electrocardiographic recording equipment did not interfere with the internal processes and overt behavior that support price formation in the experiment. Further, the graph of Figure 3 suggests that a large proportion of observations in the English institution can be explained by knowing the value of the price setter in the auction. We assess that claim in the following section.

Coefficients on value in English and Dutch auctions

We argued that the exit price of the price setter (third-person-out) of the English auction would be (at or near) that person's value. If that were the case, then knowledge of heart rate should add little to our ability to predict prices. To assess if heart rate yields any additional information beyond what has traditionally been used to explain individual behavior in the English auction, we regressed the exit price on the price setter's value and the average z score of

the price setter's heart rate during the 4 seconds preceding the exit time. For each participant we identified the auctions in which the participant was the price setter and for those auctions collected the exit price, the participant's value and mean heart rate during the four seconds prior to exiting the auction. The columns on the left of Table 1 show the regression coefficients on value and their t statistics for the data from price setters in the English auctions. The coefficients on value are significant at the .05 level for all but one of the participants. The coefficients are within 0.1 of 1.0 for 20 of the 24 participants. We interpret this result to reveal an instance in which received economic theory, a dominant strategy equilibrium, is consistent with the observed data.

To assess the degree to which knowledge of heart rate adds to our ability to predict prices in the Dutch auction, we regressed the entry value of the winner (the price setter in the Dutch auction) on the winner's value and z-score for heart rate during the four seconds immediately prior to entering the auction. By working with a subset of the observations in this auction (assuming constant relative risk aversion for the participant), we can derive an estimate of the risk coefficient using the value coefficient for the participant (Cox, Smith and Walker, 1982). Risk coefficient estimates less than one indicate risk aversion. The columns on the right of Table 1 show the risk coefficients and their significance levels from the regressions on value and heart rate for price setters in the Dutch auctions. Elements of the participants' behavior were generally consistent with the assumption of risk aversion. The estimates of the risk coefficient are significant at the .05 level for all participants. We now turn to the analysis of heart rate and its influence on our ability to predict outcomes in these auctions.

Coefficients on heart rate in English and Dutch auctions

As discussed in the previous section, we regressed exit (entry) prices against value and heart rate in the English (Dutch) auctions. Table 2 lists the t-statistics of the coefficients on heart rate and their significance levels. Figure 5 is a histogram showing the distribution of the t-statistics on heart rate. The solid bars show the distribution for the English auctions and the open bars show the distribution for the Dutch.

The t-statistics for heart rate prior to the entering the Dutch auction have a strong tendency to have a higher level of significance than the t-statistics for heart rate prior to the exiting the

English auction. Comparison of the cumulative distributions using the Kolmogorov 2 sample statistic reveals that the tendency is significant at the 5% level. We argue that this is evidence that emotional factors, as represented by instantaneous heart rate, differentially affect our ability to predict the price across these auction institutions. And thus knowledge of emotions differentially, and in a fundamental way, affects the ability to predict behavior. This differential tells us that institutions matter.

This result holds through time and is robust to different constructions of statistical tests as well as different specifications of statistical models. Furthermore, a binomial test addresses the likelihood of obtaining the observed significance levels in the Dutch auctions (under the null hypothesis that heart rate does not affect price). The likelihood of obtaining 7 or more at the 10.0% level is less than .008.

These basic findings are also robust when an interaction term is included in the regression, when we apply a log transform to value, and when we look at the direct correlation between heart rate and exit (entry) price. Furthermore the effect does not diminish if we constrain our observations to the last twenty auctions. Knowledge of heart rate enhances our ability to predict the basic behavioral outcome in the Dutch auction but does not help us predict prices in the English auction.

The cardiac system as an economic information system

Having found that heart rate has a differential influence on predictive power, we turn to the issue of why this result comes about. We suggest that the cardiac system is itself an economic information processing system that is constantly being influenced by signals from the brain about changes in the environment. In short, we view heart rate as a carrier of (economic) information.

To elaborate, economic signals arise from the environment (in this experiment the computer screen) in the form of values, the current price in the auction, and the exit or entry of other participants. This information hits the retina, is transmitted via the thalamus to the occipital lobe where it is reconstituted and sent to the frontal lobe and also back to (different nuclei in) the thalamus. From there the information goes to the hypothalamus which then excites

the vagus and secretes hormones that orchestrate the behavior of the heart. Thus in principle, the heart rate can change as the economic environment changes.

In our study, the participants received private economic information at the start and end of every auction: they were endowed with their values at the start of an auction and their profits were (not) updated upon winning (not winning) an auction. The time when auctions ended was public information for all participants as was the time when participants dropped out of the English auctions. To examine systematic variations in the participants' heart rates in response to these economic events, we graph and compare the mean heart rates during the four seconds prior to an event and the four seconds after an event. We use individually-determined z-scores to aggregate results across participants. The mean z-scores of heart rate are grouped by the participants' values (e.g., values above or below the median amount, $1.25). They are aggregated across all events of a specific type (e.g., the endowment with a value) and across all auctions of a given structure (Dutch or English). Because values are randomly assigned to participants across auctions, every participant is represented several times within each set of results.

English auctions

Figure 6 shows the mean z-scores of heart rate in the four seconds immediately before and after the release of economic information in the English auctions. The three pieces of information are the receipt of value information at the beginning of the auction (private information), the time when the second participant exits the auction (public information), and the time of the end of the auction when the third participant drops out (public information) and the remaining participant wins. The upper panel shows the change in mean heart rate for all participants with values greater than or equal to $1.25, the midpoint of the uniform distribution from which values are randomly assigned. The lower panel shows the change in mean heart rate for participants with values less than $1.25..

Heart rates in the English auction were generally below the global mean for the entire experiment. The relatively low values are unlikely to reflect a sequence effect (the block of 60 English auctions followed the block of 60 Dutch auctions) because linear trends in the observed time series of heart beats have been removed on a participant-by-participant basis. Note the z-

score of zero can be thought of as corresponding roughly to the overall mean heart rate across participants of approximately 85 beats per minute.

The upper panel of Figure 6 shows the change in mean heart rate for participants with values of $1.25 or greater. The drop in heart rate at the beginning of the auctions is marginally significant, $p < .074$ ($df = 674$) using the Wilcoxon ranked sums test. It may reveal relaxation and a temporary 'switching off' - an adaptive response. There is no significant response to information that the second person has dropped out or to the end of the auction for participants with relatively high values.

The mean heart rate for participants with values less than $1.25 is shown in the lower panel of Figure 6. For these participants, heart rates remain below the global mean throughout the English auctions. There are no significant responses to the receipt of a low value and to information that the second person has dropped out. The increase in heart rate at the end of the auctions to near the global mean is significant, $p < .046$ ($df = 655$) using the Wilcoxon ranked sums test. This finding suggests that participants with relatively low values maintain a uniform indifference or detachment as the auctions proceed but become aroused at the end of an uninteresting auction when they will soon receive a new value (at the beginning of the next auction).

Dutch auctions

Figure 7 shows the mean z-scores of heart rate in the four seconds immediately before and after (1) the beginning of the Dutch auction when participants receive a value and (2) the end of the Dutch auction when one of the participants wins the auction. The upper panel shows the mean heart rates for participants with values greater than $1.25 and the lower panel the mean heart rates for participants with less than $1.25.. In marked contrast with the English auctions, Figure 6, heart rates during the Dutch auctions, Figure 7, were often above the global mean.

The upper panel of Figure 7 shows that the mean heart rate for participants with values above $1.25 is well above the global mean at the beginning of the auction but falls below the mean before the end of the auction. It drops precipitously at the end of the auction. The drop in heart rate at the end of the Dutch auctions is significant, $p < 0.011$ ($df = 588$) using the Wilcoxon

ranked sums test. These data suggest (1) participants of Dutch auctions who received high values experienced the elevated heart rates associated with the onset of sympathetic dominance, and (2) winning a Dutch auction was marked by a sharp decrease in heart rate, a pattern consistent with a release from sympathetic dominance.

A post-hoc analysis found that winners are primarily responsible for the sharp drop in mean heart rate at the end of the auction. The drop in heart rate for winners, not shown, is significant at $p < 0.020$ ($df = 292$) while the drop for participants with the third highest value is not significant, $p > 0.20$ ($df = 295$). Only winners appear to experience the release from sympathetic dominance.

The lower panel of Figure 7 shows that the mean heart rate for participants with low values jumps upon receiving those low values and then drops below the mean before the end of the auction. The jump in heart rate was unexpected but significant, $p < 0.001$ ($df = 587$) using the Wilcoxon ranked sums test. It is experienced by participants with both the lowest and second highest values at $p < 0.007$ (with $df = 291$ and 295 respectively). The surge in mean heart rate may reveal a flash of irritation at receiving a substandard value that thwarts an opportunity for profit. It appears that nobody likes to receive a bad value in the Dutch auction.

A post-hoc analysis of the two statistically significant heart rate responses shown in Figure 7 looked for time-dependent patterns in that might be indicative of accommodation to the structure of the Dutch auction. The analysis subdivided the 60 auctions into three blocks of 20 (auctions 1-20, 21- 40, and 41-60) and conducted separate Wilcoxon tests for each of the three blocks. Figure 8 presents the data. The horizontal axis represents the three blocks of time. The upper panel shows the drop in heart rate for participants with high value at the end of the auction when one of them wins. The lower panel shows the surge in heart rate at the beginning of the auction for participants who receive low values.

The upper panel of Figure 8 shows that the sharp drop in heart rate at the end of the auctions seen in the upper panel of Figure 8 is limited to the first 20 auctions. The Wilcoxon ranked sums test reveals that the drop in heart rate at the end of the auctions is highly significant during the first block of 20 auctions, $p < 0.003$ ($df = 193$). The smaller drops during the later blocks are not significant. We interpret this result to indicate that sympathetic withdrawal upon

winning a Dutch auction is pronounced at the beginning of the session and becomes muted as time goes on. This trend suggests that participants-as-winners adapt to the structure of the Dutch auction.

The opposite pattern is seen in the bottom panel of Figure 8. The surge in heart rate when receiving a relatively low value in the Dutch auction does not become evident until the second and third blocks of time. The p-value for the first block of time is not significant, $p > .26$, but is significant thereafter, $p < .001$ for the second block and $p < .003$ for the third block. These data suggest that our participants learned to appreciate the negative implications of receiving a low value as they gained experience. We interpret the surge in heart rate to reflect irritation at the prospect of a bad outcome, i.e., not getting the chance to obtain a profit. Their inferred irritation is not a reaction to a bad outcome but rather is anticipation of that bad outcome. We interpret this result to reveal an adaptive, visceral response to economic information. This response is not evident in overt behavior but can be uncovered by observing the cardiac information processing system.

Conclusion

This is a first study to contrast the relationship between a psychophysical proxy for emotion and the choice behavior of individuals across two well-known market institutions, the English and Dutch clock auctions. Like other received knowledge from experimental economics, we have found that institutions do, in fact, matter. The economic behavior of individuals in the English-clock auction is less mediated by emotional factors than it is in the Dutch-clock auction.

Such a result should be interesting to economists for at least two reasons. First, economists are generally interested in price. Emotion is not usually considered a determinant of price In this study we adopt heart rate as a proxy for emotion to take a first step toward relating observed price to emotional factors. We have shown that this effort has promise. Second, there are clear policy implications of this type of research. A. Damasio and his colleagues have argued that, depending upon context, emotions can either augment or detract from decision making. Given this contextual dependence, there may be times in which it is desirable to choose an economic institution based on the emotional impact it has on the participants' pricing behavior. For

example, the English-clock auction may be much less affected by emotions than a similar increasing-price auction with a human auctioneer even though, from the standpoint of received economic knowledge, individual's behavior in these auctions should be the same.

Of course, given our work to date, we can only say that our findings are consistent with the hypothesis that emotions are proxied by changes in heart rate in these institutions and that heart rate in one institution provides additional explanatory power beyond typical economic variables.

Much additional work can be done. First, it is desirable to attempt to find a converging and discriminating set of measures of emotion. Heart rate is just one metric; others include galvanic skin response and subjective verbal reports. Until we perform studies using more discriminating metrics it will be hard to say precisely which emotions are at work and if, in fact, we are actually observing a truly emotional response.

There are a wide variety of institutions that are also available for study. How does a double auction designed to improve bid and ask offers compare with the two auctions we have already studied? Are emotional factors dampened or increased by the presence of a human auctioneer, by public bidding, or by casual observers? Can we elaborate the determinants of contextual sensitivity to uncover when and why institutions matter?

Typically the phrase "Institutions Matter" has been used by economists to apply to wider societal issues. The rules and means governing economic transactions have been used to assess the relationship between property rights and the performance of a society (North,1994), the relationship between forms of deposit insurance and the existence of moral hazard (Allen, 2001), as well as the interplay between a well-functioning economy, its fairness, and corruption (Broadman and Recantini, 2001). Such institutions may also not be immune to having emotions play a role.

References

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. Human Factors **19** (2), 111-170.

Allen, F. (2001). Do financial institutions matter? Journal of Finance **56** (4), 1165-1175.

Backs, R. W. (1995). Going beyond heart rate: Autonomic space and cardiovascular assessment of mental workload. The International Journal of Aviation Psychology **5** (1), 25-48.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. Cognition **50** (1-3), 7-15.

Berg, J., Dickhaut, J, and McCabe, K. (in review). Risk preference instability across institutions: A dilemma. Proceedings of the National Academy of Sciences.

Bechara, A., Damasio, H., Damasio, A. R., & Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. Journal of Neuroscience **19**, 5473-5481.

Bechara, A., Damasio, H., Tranel, D., & Damasio A. R. (1997). Deciding advantageously before knowing the advantageous strategy. Science **275** (5304), 1293-1295.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. Cerebral Cortex **6**, 215-225.

Broadman, H. G., & Recantini, F. (2001). Seeds of corruption: Do market institutions matter? Washington, DC: The World Bank.

Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An $H_2{}^{15}O$ PET study of Stroop task performance. Neuroimage **2**, 264-272.

Collet, C., Vernet-Maury, E., Delhomme, G., & Dittmar, A. (1997). Autonomic nervous system response patterns specificity to basic emotions. Journal of the Autonomic Nervous System **62**, 45-57.

Cox, J. C., Roberson, B., & Smith, V. L. (1982). Theory and behavior of single object auctions. Research in Experimental Economics **2**, 1-43

Damasio, A. R. (1994). Descartes' error: Emotion, reason, and the human brain. New York: Putnam.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences **351**, 1413-1420.

Damasio, A. R. (1999) The feeling of what happens. Orlando: Harcourt Brace.

Dickhaut, J., McCabe, K., Nagode, J. C., Smith, K., & Pardo, J. (2003). The role of context in choice: The effect of different comparison gambles on risky choice. Proceedings of the National Academy of Science **100** (6), 3536-3541.

Dickhaut, J., Smith, K., McCabe, K., Peck, N. R., & Rajan, V. (2001). Inferred Brain Function in the Formation of Allocations in an English Auction. In R. Zwick (Ed.) Experimental Business Research. Hong Kong: Hong Kong University of Science and Technology Press.

Dolan, R. J., Fletcher, P., Morris, J., Kapur, N., Deakin, J. F. W., & Frith, C. D. (1996). Neural activation during covert processing of positive emotional facial expressions. Neuroimage **4**, 194-200.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. Science **221**, 1208-1210.

Glimcher, P. W. (2002). Decisions, uncertainty, and the brain: The science of neuroeconomics. Cambridge, MA: MIT Bradford books.

Isaac, R. M., & Jackson, D. (1994). Just who are you calling risk averse? Journal of Risk and Uncertainty **20** (2), 177-187.

Kagel, J. & A. Roth, A. (1995). Handbook of Experimental Economics. Princeton, NJ: Princeton University Press.

Lane, R. D., & Schwartz, G. E. (1990). The neuropsychophysiology of emotion. Functional Neurology **5** (3), 263-266.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. Psychophysiology **27** (4), 363-384.

Lo, A., & Repin, D. (2002). The psychophysiology of real-time financial risk processing. Journal of Cognitive Neuroscience **14** (3), 323-339.

Loewenstein, G., (1996). Out of control: Visceral influences on behavior. Organizational Behavior and Human Information Processes **65**(3), 272-292.

Marshuetz, C., Smith, E. E., Jonides, J., DeGutis, J., & Chenevert, T. L. (2000). Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. Journal of Cognitive Neuroscience **12** (6, Supp. 2), 130-144.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. Proceedings of the National Academy of Science **98** (20), 11832-11835.

McCabe, K., Rassenti, S., & Smith, V. (1990). Auction institutional design: Theory and behavior of simultaneous multiple-unit generalizations of the Dutch and English auctions. The American Economic Review **80**, 1276-1283.

McCabe, K., Rassenti, S., & Smith, V. (1991). Testing Vickrey's and other simultaneous multiple unit versions of the English auction. In R. Mark Isaac (Ed.) Research in Experimental Economics **4**, 45-79. Greenwich, CN: JAI Press.

Mellers, B., Schwartz, A., Ho K., & Ritov, I (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. Psychological Science **8**(6), 423-429.

Mellers, B., Schwartz, A., Ho K., & Ritov, I (1999). Emotion-based choice. Journal of Experimental Psychology General **128**, 332-345.

Menon, V., Rivera, S. M., White, C. D., Glover, G. H., & Reiss, A. L. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. Neuroimage **12**(4), 357-365.

North, D. C. (1994). Institutions matter. Economics history working paper 9411004. Economics working paper archive. St Louis: Washington University.

Pardo, J. V., Pardo, P.J., Janer, K. W., & Raichle, M. E. (1990). The cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. Proceedings of the National Academy of Science **87**, 256-259.

Plott, Charles R. (1982) Industrial organization theory and experimental economics. Journal of Economic Literature **20**, 1485-1527.

Rogers, R. D., Everitt, B. J., Baldacchino, A., Blackshaw, A. J., Swainson, R., Wynne, K., Baker, N. B., Hunter, J., Carthy, T., Booker, E., London, M., Deakin, J. F. W., Sahakian, B. J., & Robbins, T. W. (1999). Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: Evidence for monoaminergic mechanisms. Neuropsychopharmacology **20** (4), 322-339.

Rustichini, A., Dickhaut, J., Ghirardato, P., Smith, K., & Pardo, J. (this volume). A brain imaging study of procedural choice Games and Economic Behavior.

Sinha, R., Lovallo, W. R., & Parsons, O. A. (1992). Cardiovascular differentiation of emotions. Psychosomatic Medicine **54**, 422-435.

Slovic, P., Finucane, M. L.,. Peters, E.,, & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich et al. (Eds.), Intuitive Judgment: Heuristics and Biases. Cambridge: Cambridge University Press.

Smith, K., Dickhaut, J., McCabe, K., & Pardo, J. (2002). Neuronal substrates for choice under ambiguity, risk, gains, & losses. Management Science **48** (6), 711-718.

Smith, V.L. 1982. Microeconomic Systems as an Experimental Science. American Economic Review **72** (5). 923-55.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Psychology **18** (6), 643- 662.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed-bid tenders. Journal of Finance **16**, 8-37.

Tables

Table 1. Estimated value coefficients in the English auctions and risk coefficients in the Dutch auctions.

English			Dutch	
Value coef.	P	Player	Risk coef.	P
1.004	**.000**	1	0.692	**.000**
1.055	**.000**	2	0.750	**.000**
1.002	**.000**	3	0.713	**.000**
0.654	**.000**	4	0.465	**.014**
0.862	**.000**	5	0.741	**.000**
1.002	**.000**	6	0.744	**.000**
0.916	**.000**	7	0.748	**.000**
1.005	**.000**	8	0.827	**.000**
0.997	**.000**	9	0.986	**.000**
-0.159	**.289**	10	0.865	**.000**
0.997	**.000**	11	0.730	**.000**
0.402	**.007**	12	0.670	**.000**
0.940	**.000**	13	0.865	**.000**
1.007	**.000**	14	0.893	**.000**
0.983	**.000**	15	0.860	**.000**
1.007	**.000**	16	1.005	**.000**
0.599	**.000**	17	0.816	**.005**
1.005	**.000**	18	0.932	**.000**
1.002	**.000**	19	0.901	**.000**
1.006	**.000**	20	0.948	**.000**
1.012	**.000**	21	0.917	**.000**
1.022	**.000**	22	0.706	**.012**
1.001	**.000**	23	0.663	**.000**
0.944	**.000**	24	0.752	**.000**

Table 2. Absolute values of the t-statistics for the regression coefficients for heart rate and their significance levels in the regressions of exit (entry) price on value and heart rate in the English (Dutch) auctions.

English			Dutch	
abs(t-stat)	P	Player	abs(t-stat)	P
0.330	**.373**	1	2.354	**.014**
0.015	**.494**	2	0.963	**.186**
0.255	**.402**	3	1.504	**.077**
0.397	**.348**	4	1.001	**.173**
1.156	**.131**	5	0.285	**.389**
0.619	**.279**	6	1.177	**.132**
0.447	**.331**	7	1.772	**.052**
0.293	**.389**	8	0.997	**.171**
1.448	**.122**	9	0.646	**.266**
1.220	**.121**	10	0.621	**.277**
1.427	**.124**	11	1.228	**.130**
1.266	**.121**	12	1.507	**.075**
0.009	**.497**	13	0.656	**.263**
0.126	**.451**	14	0.911	**.189**
0.008	**.497**	15	0.895	**.197**
0.604	**.279**	16	1.015	**.163**
1.136	**.133**	17	0.701	**.261**
0.714	**.253**	18	1.861	**.042**
0.314	**.380**	19	2.149	**.023**
0.875	**.204**	20	0.444	**.332**
1.153	**.136**	21	1.308	**.105**
0.287	**.391**	22	2.517	**.043**
0.810	**.219**	23	1.239	**.117**
0.543	**.299**	24	0.143	**.444**

Figure captions

Figure 1. Heart-rate response to induced emotions (solid bars from Ekman, Levenson, & Friedman, 1983; open bars from Sinha, Lovallo, & Parsons, 1992).

Figure 2. Computer display for the experimental auctions.

Figure 3. The difference between the value of the last person to exit and the exit price in the English auctions. The value of the third person out appears to play a pivotal role in price determination, replicating known results without heart rate equipment.

Figure 4. Observed entry prices plotted against value in the Dutch auctions. The thicker, lower line represents the risk neutral bid for a four-person auction. The graph appears to indicate risk averse behavior since the group of subjects tends to pay a premium over expected value to win the auction.

Figure 5. Comparison of the distributions of the absolute values of the t-statistics for heart rate in the regressions of exit (entry) price on value and heart rate in the English (Dutch) auctions. The distribution for Dutch is significantly shifted to the right.

Figure 6. Change in the average z-score of heart rate across events in the English auctions: upon receiving the private value, when the second player exits, and at the exit price. The upper panel shows the heart rates for all players with values greater than or equal to $1.25, the midpoint of the uniform distribution from which values were randomly assigned. The lower panel shows heart rates for players with values less than $1.25.

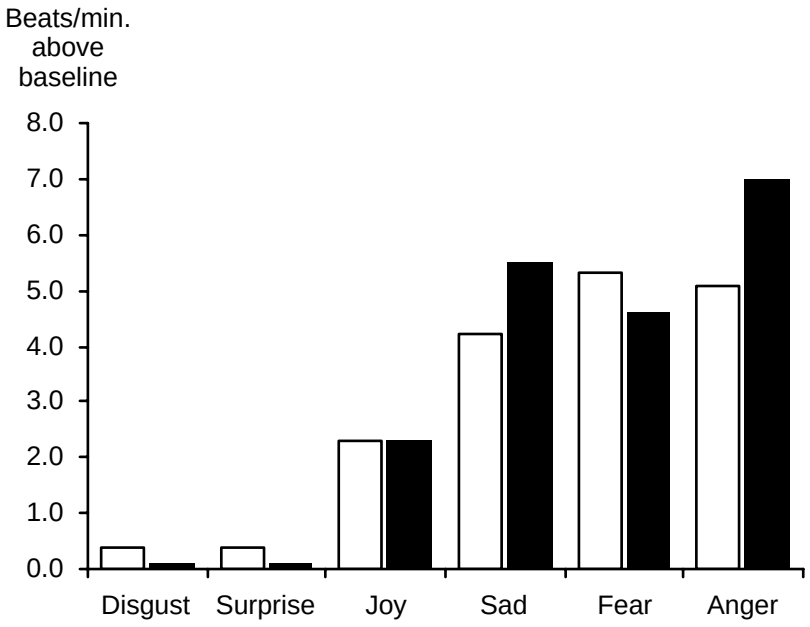
Figure 7. Change in the average heart rate across events in the Dutch auctions: upon receiving the private value and at the end of the auction. The upper panel shows heart rates for players with the values greater than $1.25. The lower panel shows heart rates for players values less than $1.25.

Figure 8. The time-dependency of the two significant changes in heart rate in response to the release of economic information in the Dutch auctions. The 60 auctions were subdivided into three blocks of 20. The upper panel shows the change in average heart rate for players with the

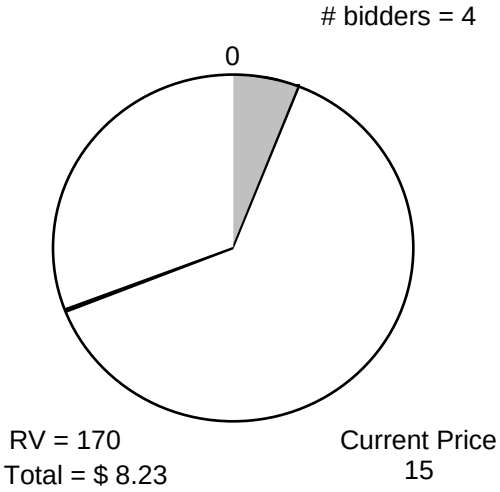
values above $1.25 at the end of the auction, when one of them wins. The lower panel shows the change in heart rate for players as they receive the values below $1.25 at the beginning of the auction.

Figures

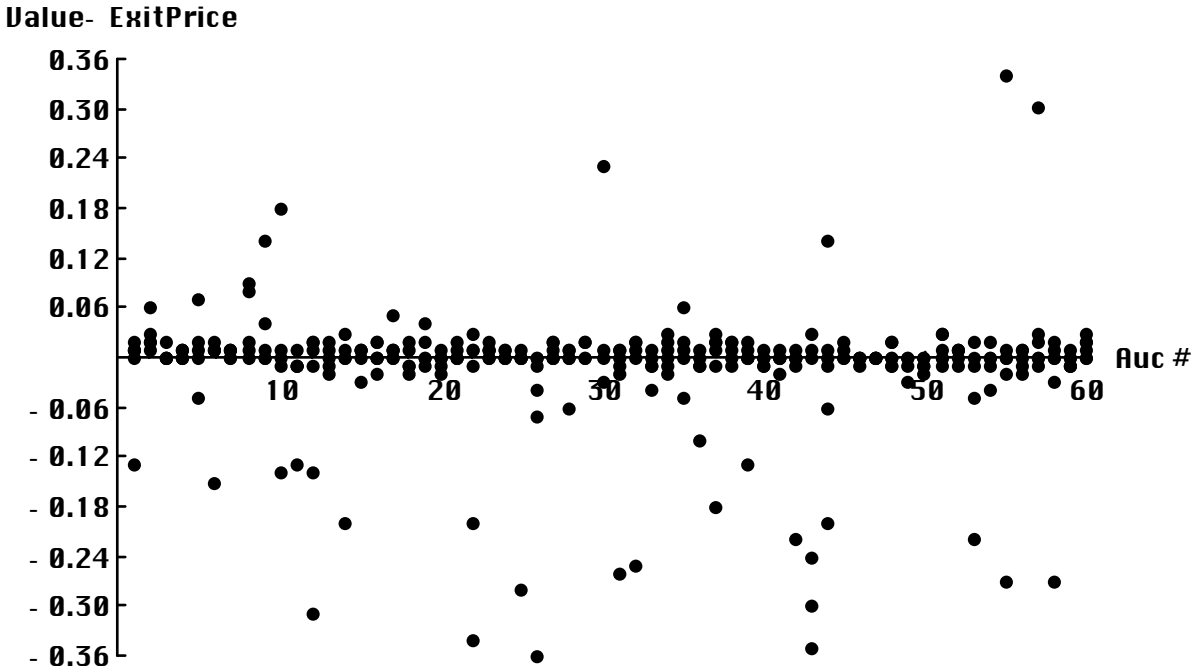
1



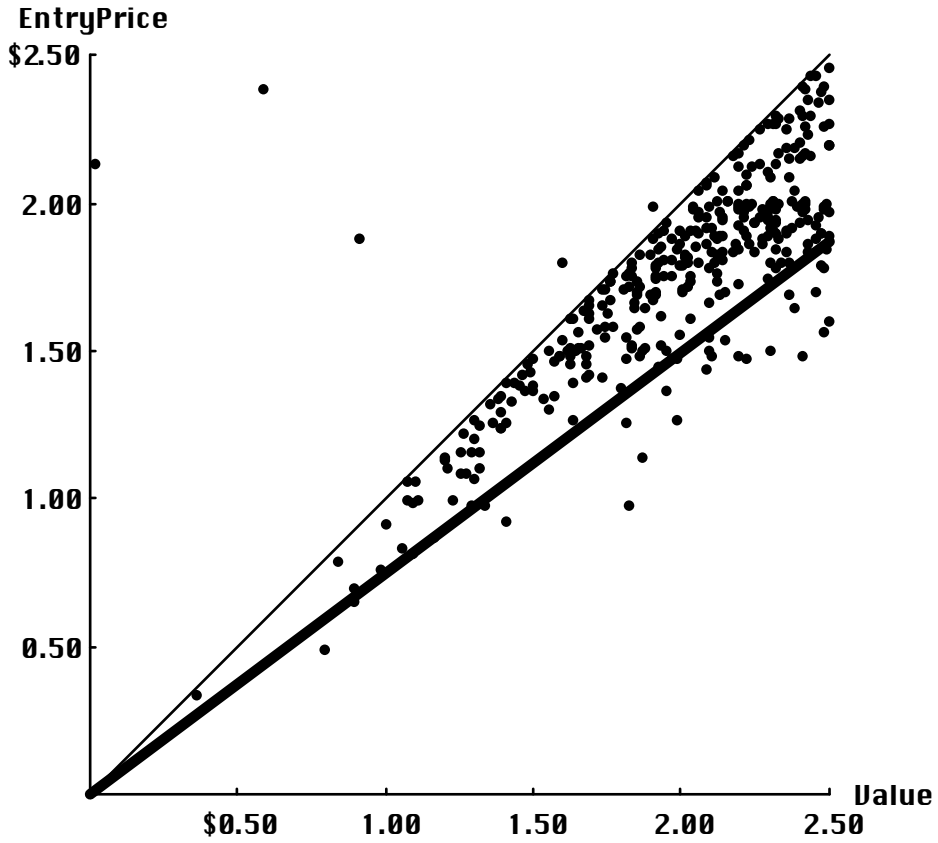
2


bidders = 4
0
RV = 170
Total = $ 8.23
Current Price
15

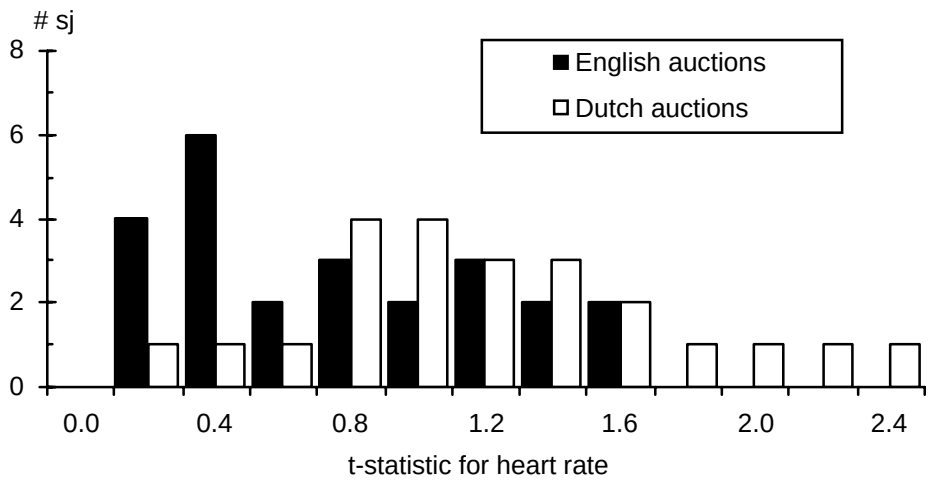
3


Value- ExitPrice
0.36
0.30
0.24
0.18
0.12
0.06
- 0.06
- 0.12
- 0.18
- 0.24
- 0.30
- 0.36
10
20
30
40
50
60
Auc #

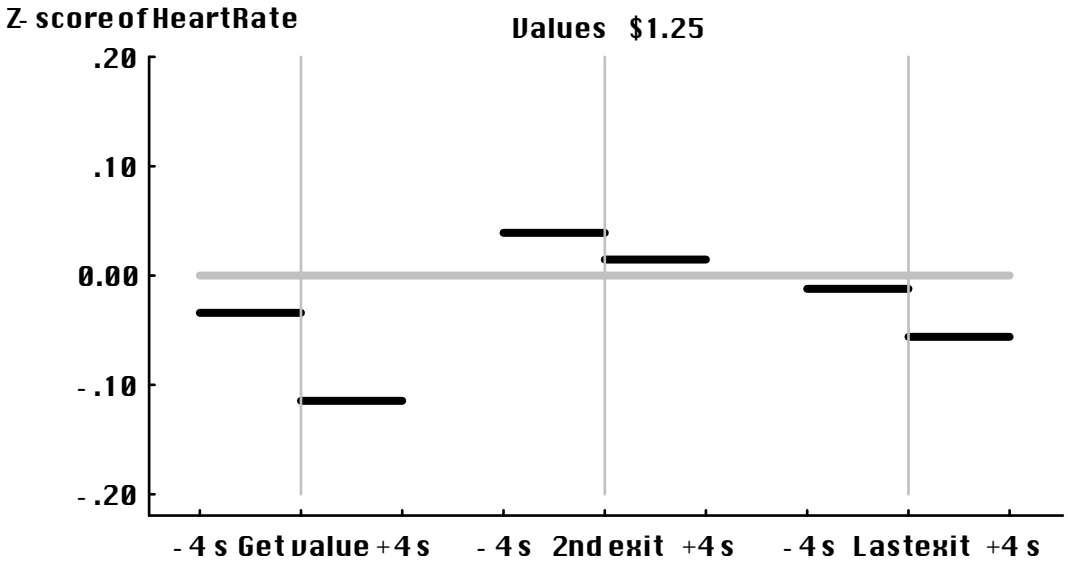
4

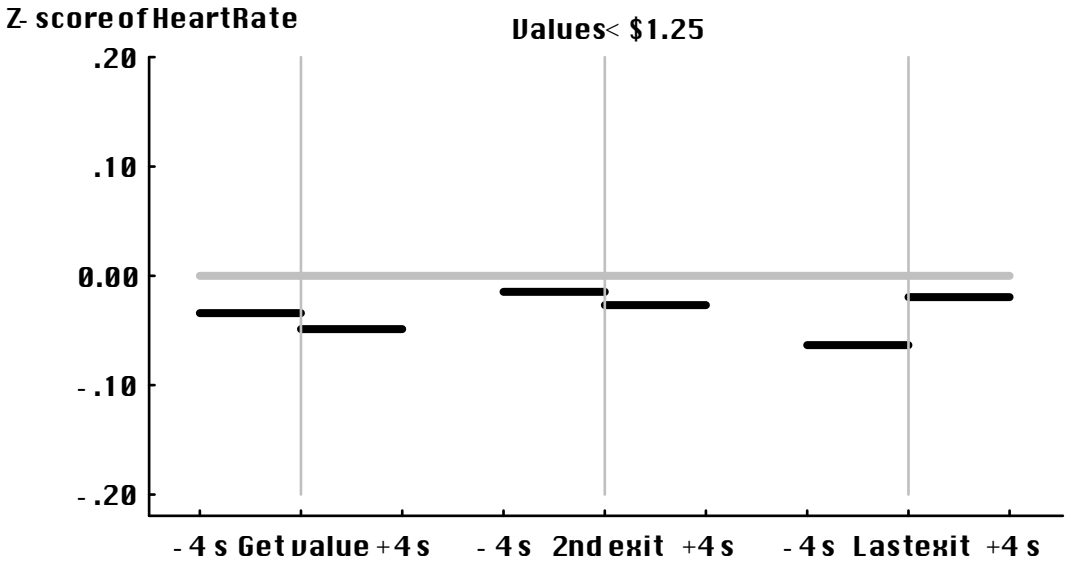


5

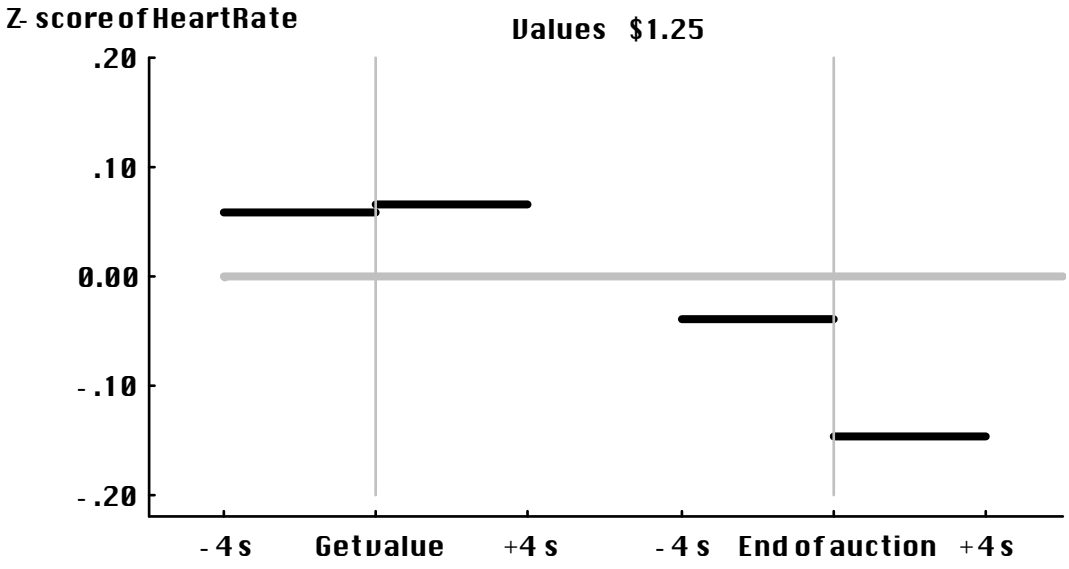

sj
8
6
4
2
0
English auctions
Dutch auctions
0.0
0.4
0.8
1.2
1.6
2.0
2.4
t-statistic for heart rate

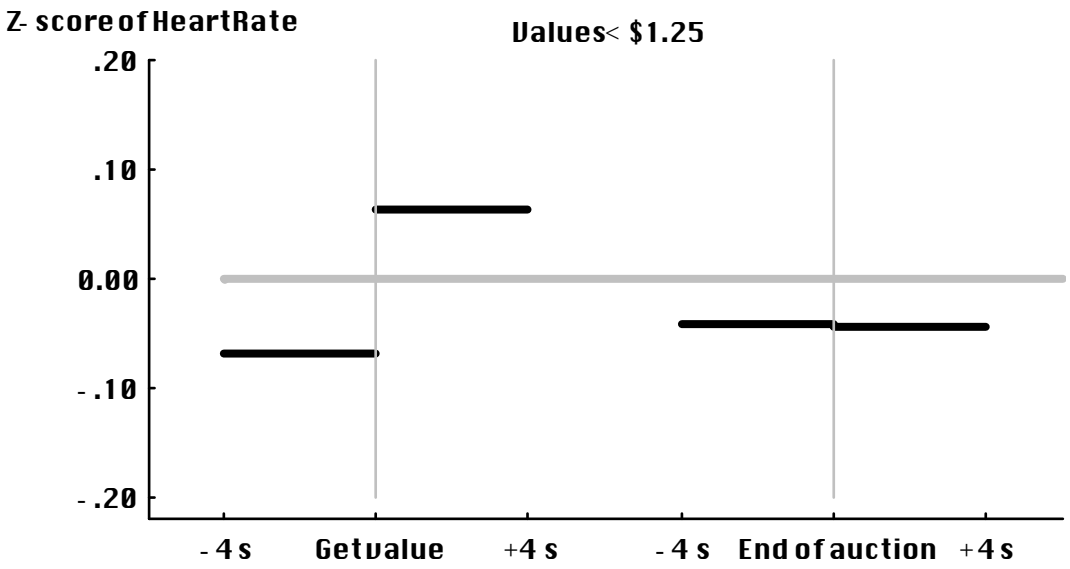
6


Z- score of HeartRate
Values $1.25
.20
.10
0.00
- .10
- .20
- 4 s Get value +4 s
- 4 s 2nd exit +4 s
- 4 s Last exit +4 s

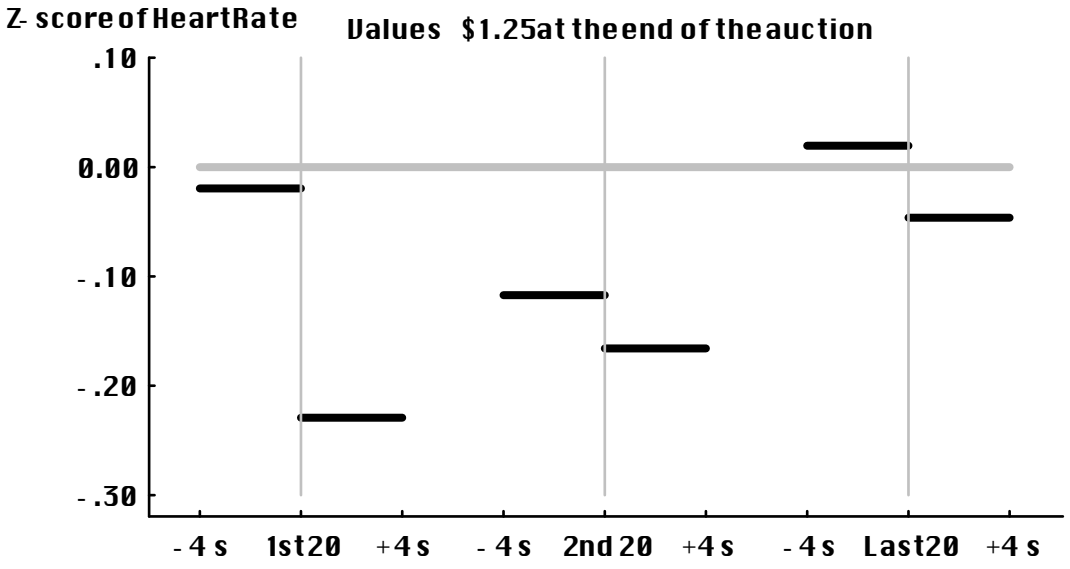

Z-scoreofHeartRate
Values< $1.25
.20
.10
0.00
-.10
-.20
-4s Getvalue +4s
-4s 2ndexit +4s
-4s Lastexit +4s

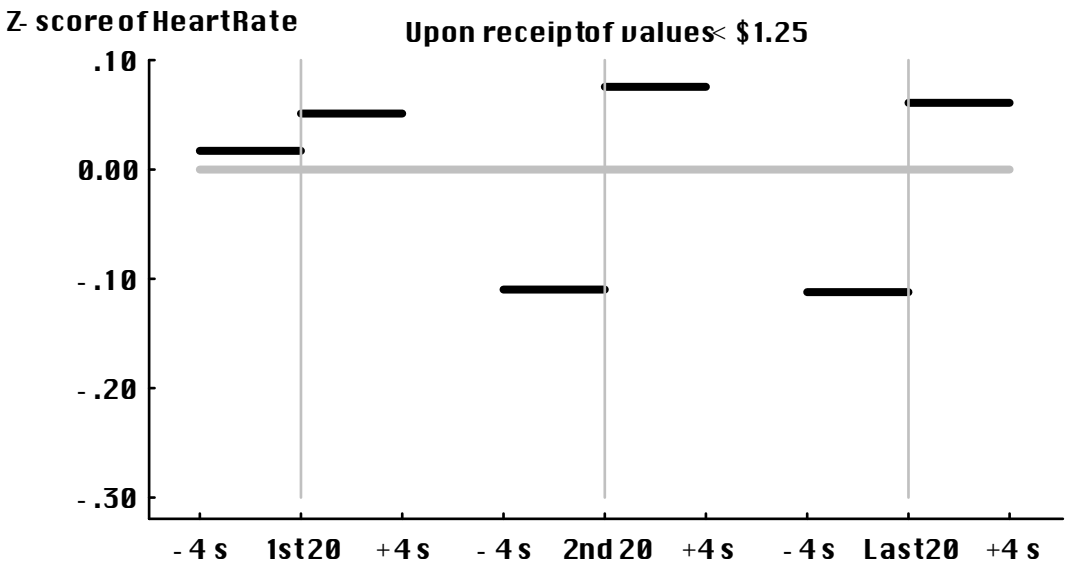
7


Z-scoreofHeartRate
Values $1.25
.20
.10
0.00
-.10
-.20
-4s
Getvalue
+4s
-4s
Endofauction
+4s


Z-score of Heart Rate
Values < $1.25
.20
.10
0.00
-.10
-.20
-4 s
Get value
+4 s
-4 s
End of auction
+4 s

8


Z-score of Heart Rate
Values $1.25 at the end of the auction
.10
0.00
-.10
-.20
-.30
-4 s
1st 20
+4 s
-4 s
2nd 20
+4 s
-4 s
Last 20
+4 s


Z- score of Heart Rate
Upon receipt of values < $1.25
.10
0.00
- .10
- .20
- .30
- 4 s
1st 20
+4 s
- 4 s
2nd 20
+4 s
- 4 s
Last 20
+4 s

Linköping Institute of Technology Dnr IKP 214/03-23
Progress Report covering the period 1 July 2005 – 30 September 2005
MURI OPUS -- DAAD19-01-1-0621

Principal Investigator: Kip Smith, Ph.D.
Division of Industrial Ergonomics

Executive Summary

Two allied lines of research were pursued during the third quarter of 2005 as part of the MURI-OPUS project at the Linköping Institute of Technology. (1) We continued the new series of laboratory experiments on collaborative, distributed command and control that was discussed in detail in the second quarter 2005 report. (2) We obtained access to Swedish Army recruits as participants in our long-standing program of experiments on the impact of remote command and control using the paintball paradigm. In addition, we presented MURI-OPUS-sponsored research at the 49th Annual Meeting of the Human Factors and Ergonomic Society. On the way to the meeting, a portable PC bought by the MURI-OPUS grant and owned by the University of Central Florida was misplaced by the Transportation Security Administration at O'Hare Airport. The computer appears to be irretrievably lost. This memo discusses these developments and the presentation.

Laboratory Experiments on Collaborative Distributed Command and Control

The purpose of the laboratory experiments is to investigate cultural differences in norms for decision making in collaborative, distributed command and control. The experimental platform is the C3Fire microworld. During the third quarter we conducted experiments with participants drawn from the communities of Bosnian, Iranians, and Indians living in Sweden. Data collection and analysis will continue in the fourth quarter.

Here we discuss the theoretic foundation for our data collection and analyses. We begin by providing a brief overview of the pioneering work of Geert Hofstede's and explaining why the field of cross-cultural psychology "has moved on" from Hofstede. We then turn to a discussion of the theory-driven perspective on core human values offered by Shalom Schwartz (1992, 1994). The Schwartz circumplex model refines rather than contradicts Hofstede's empirical findings and provides the theoretic grounding for our research. We have adopted Schwartz's model and instrument and are using them to assess the cultural differences in norms for decision making that are likely to emerge in multination (multi-cultural) operations.

Hofstede

Hofstede (1980) conducted the pioneering work on cross-cultural variations in human values. His aim was to map cultures on the basis of variables that can be directly linked to fundamental human values. This aim remains the central theme of much of cross-cultural psychology (Bond and Smith, 1996).

His approach was (1) to collect morale surveys that asked questions about values from more than 100.000 workers for a large multinational corporation (IBM) who had been born and raised in more than 50 different countries, (2) to use factor analysis to extract core 'dimensions' of human values, and (3) to compare across countries the relative weights given to each dimension. This effort generated four core dimensions - power distance, uncertainty avoidance,

individualism/collectivism, and masculinity/femininity – and national rankings for more than 50 countries along each dimension.

Table 1 provides brief definitions of each of Hofstede's four core dimensions and presents the Hofstede's rankings for countries whose (ex-) citizens we are likely to study in our project. For example, Sweden's rankings place it near the bottom of the list for power distance, uncertainty avoidance, and masculinity and near the top of the list for individualism. The implication of these rankings is that, as a national culture, Swedes are comfortable expressing dissent, are willing to bend rules if it is in the best interest to do so, are keen on family and workgroup cohesion, and are relatively unconcerned about individual recognition. The rankings for India, Iran, Spain, and Yugoslavia are markedly different.

While informative, Hofstede's rankings are averages calculated from thousands of individuals. As averages, they reveal only the central tendency and conceal the variability in the data. The rankings falsely imply that national cultures are "unitary systems free from variation, conflict and dissent" (Smith and Bond, 1996, p. 42). There is, of course, at least as much variability within a culture as across cultures. For example, while the Swedish culture may be relatively unconcerned with exogenous rewards and recognition, there are undoubtedly a large number of Swedish individuals who seek the trappings of achievement. Thus, it is inappropriate to apply Hofstede's rankings to individuals. Hofstede himself emphasized this point: his work applies to nations and not to individuals. We cannot and should not use it to explain how individuals from different cultures will interact in our simulated command-and-control exercises.

This is, in fact, one of the main reasons why cross-cultural psychology has moved on from Hofstede: No one wants to commit what Hofstede called "the ecological fallacy" (Hofstede, 1980, p. 29) – misapplying his national-level findings at the level of individuals.

A second source of dissatisfaction with Hofstede's work is that it was purely empirical. Hofstede offered no theoretic underpinnings for the questions in his surveys or for the four core dimensions that emerged from his factor analysis. While the lack of a theoretic foundation enabled him to uncover 'core dimensions' of human values, it precludes prediction.

These considerations and conversations with many members of the community of cross-cultural psychologists (at the 9th International Conference on Cross-Cultural Psychology) led us to move on from Hofstede and to follow the lead of Shalom Schwartz.

Table 1. – Hofstede's four core dimensions of human values and his rankings of selected countries on those dimensions. A small number indicates a high ranking. The lowest ranking is 53. Definitions and data from Smith and Bond (199x, p. 45).

> Power distance: The amount of respect and deference between those in superior and subordinate positions. High rankings (low numbers) indicate the expectation of considerable deference.
>
> Uncertainty avoidance: A focus on planning and stability to deal with life's uncertainties. High rankings indicate a strong emphasis on stability.
>
> Individualism: One's identity is defined by personal choices and achievements. In contrast,

Collectivism: One's identity is defined by the character of the collective groups to which one is more or less permanently attached. High rankings reflect high individualism.

Masculinity: Emphasizes achievement. Femininity: Emphasizes interpersonal harmony. High rankings indicate an emphasis on achievement.

Country	Power distance	Uncertainty avoidance	Individualism	Masculinity
India	10	45	21	20
Iran	29	31	24	35
Spain	31	12	20	37
Sweden	47	49	10	53
Yugoslavia	12	8	34	48

Schwartz

Shalom Schwartz (1992, 1994) is the central figure in the post-Hofstede era of theoretically-grounded cross-cultural psychology. Schwartz's contributions are numerous and influential. His configural model of value types, reproduced as Figure 1, is the most widely accepted theory of human values.

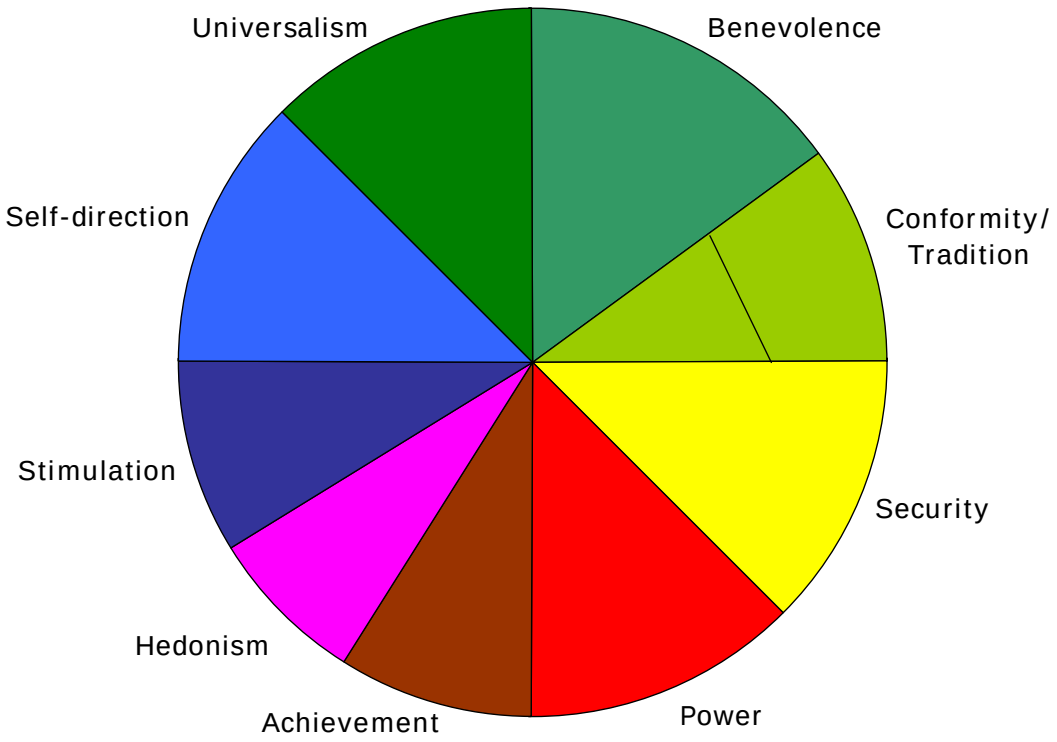


Figure 1. The configural model of the structure of core human values. (From Schwartz, 1992).

Schwartz rooted the study of values in the framework of human evolutionary needs (Schwartz and Bilsky, 1987, 1990). He identified in the previous literature a set of 56 human values, e.g.,

creativity, curiosity, pleasure, wealth, and health. Each of the 56 had been noted in more than one culture and, he argued, meets at least one evolutionary need. He advanced a dozen hypotheses that led him to propose his configural model of 10 "value types." The 56 values populate and define the configural model of value types. The circular representation emphasizes the inter-relatedness of the 10 value types and the 56 constituent values.

Schwartz developed a survey instrument which he and his colleagues used to collect data from individuals from more than 50 national cultures. One of these colleagues was Åke Daun of the University of Stockholm. Schwartz used these data to test and refine his hypotheses and the configural model of value types. Figure 1 is the product of this refinement process. He and others have used his instrument and the configural model of value types to explore and explain cross-cultural differences in a host of domains and applications. In what follows, we discuss selected aspects of his model and his instrument in greater detail.

Values and needs

Schwartz and Bilsky (1987, 1990) proposed that "values represent, in the form of conscious goals, three universal requirements of human existence to which all individuals and societies must be responsive:

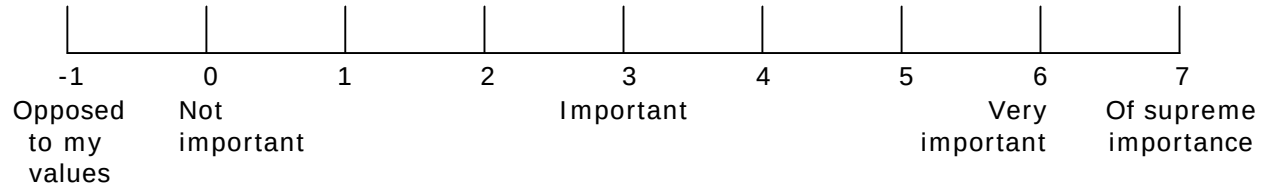
1 needs of individuals biological organisms,
2 requisites of coordinated social interaction, and
3 survival and welfare needs of groups.

From an evolutionary viewpoint (Buss, 1986), these goals have crucial survival significance." (Schwartz, 1992, p. 4).

Schwartz is making several claims here. The first is that a value (e.g., wealth or health) is a goal that individuals in all cultures pursue to one degree or another. The second is that these goals exist because they serve one or more of three basic biological or societal needs. These claims explain both the origin and universality of human values and, in the process, justify the search for cross-cultural differences in the weights individuals attach to them.

The instrument

To test these claims, Schwartz developed a survey instrument consisting of 56 questions, one for each of the values commonly cited in the literature. The survey asks respondents to rate each value "As a guiding principle in my life," using the following nine-point scale. The –1 is unusual but highly useful. It allows respondents to indicate a "negative" value – a value they seek to avoid expressing or promoting through their choices and behavior.



The values are presented in two lists: the first contains nouns (e.g., equality, freedom, excitement in life); the second contains adjectives (e.g., humble, helpful, curious). This splits

the survey into two manageable parts. Prior to rating the values on each list, respondents are instructed to read the whole list, and to chose and rate the value most important to them, and then to chose and rate the value they most oppose. This procedure serves to anchor the 9-point scale and to encourage retrospection when rating the values.

We have adopted this instrument and are adhering to Schwartz's methodology.

The configural model of 56 values and 10 value types

Schwartz developed a series of first principles of compatibility and conflict between the 56 values. These principles were rooted in the three basic biological or societal needs. For example, Schwartz argued that values that support security are universally seen as compatible with values that support the maintenance of tradition and conformity but as in conflict with values that support stimulation and individuation.

These considerations led Schwartz to propose the hypothesis that the 56 values would sort themselves into the 10 "value types" itemized in Table 2. He further proposed that these value types could be represented as a circle, with each value type forming a segment of the circle as shown in Figure 1. Adjacent value types were proposed be to be most compatible and those on opposite sides of the circle were proposed to be in most conflict. As shown in Figure 1, Schwartz hypothesized that "value types that serve primarily individual interests (power, achievement, hedonism, stimulation, self-direction) form a contiguous region opposed to another contiguous region formed by the three value types that serve primarily collective interests (benevolence, tradition, conformity)." (Schwartz, 1992, p. 13). Universalism and security serve both and are therefore located on the boundaries.

Table 2 – Schwartz's 10 value types and the goal they serve

	Value type	**Goal**
1	Self-direction	independent thought and action; autonomy and independence
2	Stimulation	the organismic need for variety; excitement, novelty, and challenge
3	Hedonism	pleasure or sensuous gratification for oneself
4	Achievement	demonstrating competence to obtain social approval; the focus is social esteem
5	Power	attainment of social status and prestige, and control or dominance over people and resources; the focus is, again, social esteem
6	Security	safety, harmony, and stability of society, of relationships, and of self
7	Conformity	self-restraint in everyday interaction; restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms
8	Tradition	respect, commitment, and acceptance of the customs and ideas that one's culture or religion impose on the individual
9	Benevolence	concern for the welfare of close others
10	Universalism	understanding, appreciation, tolerance, and protection for the welfare of all people and for nature

Schwartz administered his instrument to individuals in more than 50 countries to test his hypotheses. The hypotheses were largely supported. The "substantial match in most samples

between the observed and the hypothesized contents and structure of the value types … strongly supported all the hypotheses regarding compatibilities and conflicts between value types." (Schwartz, 1992, p. 31).

The Schwartz model has been tested innumerable times since its initial publication. With few exceptions (e.g., China), individuals in all literate cultures appear to implicitly distinguish the 10 value types when assessing the importance of specific values as guiding principles in their lives. The model represented by Figure 1 appears to be an exhaustive and near universal classification of motivational values.

Its widespread acceptance and explanatory power have led us to adopt the Schwartz model and instrument in our laboratory.

Example application of the Schwartz model

Schwartz's model provides a basis for generating hypotheses that link responses to his value questionnaire to performance measures (dependent variables) captured by C3Fire. When participants answer the questionnaire they indicate which of the 56 values are central to their value structure and which are less important. Different participants have different value structures which, ideally, correlate with differences in their patterns of behavior.

For example, if our data replicate the findings of both Hofstede and Schwartz, most Swedes will indicate that values in the value type "Universalism" are more central to their world view than values in the value type "Power." However, some Swedes will buck this trend and value Power more than Universalism. The Schwartz model uses natural variability as a source of explanatory power.

We propose that participants in our experiments who value Power more than Universalism will be more likely to take a dominating role in the C3Fire sessions than participants who value Universalism more than Power. We hypothesize that the "power people" will issue more verbal commands and move more trucks than the "universalists." We expect to see a stronger correlation between rankings on Power and the dependent variable Number-of-Trucks than between rankings on Universalism and Number-of-Trucks.

Following Schwartz, "every hypothesis that specifies the association of one value type with an outside variable (e.g., a C3Fire performance measure) has clear implications for the associations of the other values types as well." (Schwartz, 1992, p. 54). The strong correlation with Power should also be expressed as a similarly strong correlation with value types that are adjacent to Power in the value structure (e.g., Achievement). If our hypothesis holds, the correlations should decrease monotonically to a minimum at Universalism. Accordingly, our hypothesis about Power, Universalism and Number-of-Trucks can be represented by the curve shown in Figure 2. The horizontal axis represents the 10 value types. The vertical axis is the strength of correlation increasing upwards.

By combining Schwartz's abstract level of measurement with our contextually specific measures from C3Fire, we expect to be able to generate and test a myriad of hypotheses about values structures and performance variables in our simulated command and control center.

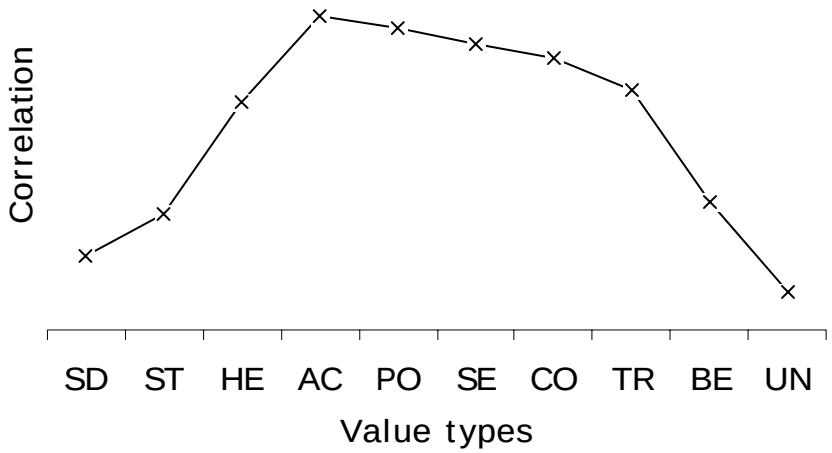


Figure 2 The hypothesized pattern of correlations between value types and the C3Fire variable Number-of-Trucks

Presentation at HFES

We have been conducting field experiments on the impact of remote command and control on soldier cognitive performance using the paintball paradigm since the inception of the MURI-OPUS project. We have documented that participants take significantly longer to obey a verbal command to move or to shoot when that command is issued by a remote, rather than a collocated, leader. In 2004 we extended this line of work to investigate whether cutaneous communication in the form of vibro-tactile stimulation might compensate, in part, for this decrement performance. Our results suggested that it can. During this quarter, my student, Martin Liberg, and I presented these results in a lecture presentation at the 49th Annual Meeting of the Human Factors and Ergonomic Society in Orlando, Florida. The paper, entitled Paintball 2: A study of remote-controlled vibro-tactile stimulation for the simulation of leader presence, was attached to the second quarter 2005 report. The response to the presentation was enthusiastic.

Lost Computer

The following is the text of an email sent to Barry Wick of UCF on 19 September 2005.

Barry

Kip Smith here.

I am writing you to inform you about the loss of UCF property. The computer was purchased nearly 5 years ago as part of the MURI grant that Peter Hancock leads. The circumstances of the loss are described in the email below to United Airlines.

I have contacted both the Transportation Safety Misadministration and United and heard nothing in return.

Because the computer has a UCF sticker prominently displayed, there is a remote chance that someone at UCF will hear directly from the TSA or airline. There is no way to connect me to the computer without probing the files.

The computer is old. Some parts don't work anymore. So it is not a big loss. However, it is lost and you or someone at UCF need to know. Sorry about this.

Kip Smith

Dear United baggage people at O'Hare,

My name is Christian Smith.

I think you or the TSA may have my old Dell Inspiron laptop. This black PC was in a black carrying case with no strap. It was in my checked luggage when I checked in for United Flight 665 from ORD to Minneapolis on Tuesday Sept 20.

The laptop has two distinctive features:

1 It is owned by the University of Central Florida and has a little yellow sticker saying so.

2 It has an empty battery slot on the right side in the front.

I checked my bag containing the laptop with the desk agent at approximately 8:30 PM after wheeling it from the International Terminal. My flight from Copenhagen to ORD was 2.5 hours late so I got rebooked onto flight 665 from ORD to MSP and had to wheel the bag myself on the tram and to Terminal 1.

I KNOW the PC was in the bag when I checked in for Flight 665.

Flight 665 arrived at MSP after midnight. I was exhausted. My boy grabbed my bag off the belt. I was not aware that the PC was missing until the next morning. When I opened the bag the PC was gone. In its place was one of those forms from the Transportation Security Administration.

I called you (1 733 601 3295) and left a rambling message Wednesday morning giving you my FAX number (507) 389 5625.

What I think happened is that the TSA people at ORD scanned the bag, saw there was a PC inside it, thought that suspicious, took the PC out of the bag, closed the bag and sent the bag on its way. They ended up with a PC and no way of knowing where to put it. So I think it is in the lost and found somewhere in the bowels of O'Hare, probably with United.

Field Experiment with Soldiers as Participants

Upon my arrival in Sweden in the summer of 2003, I contracted Berndt Brehmer of the Swedish Defense College to lay the foundation for collaboration. This effort has paid off. The microworld used as the experimental platform for the experiments discussed above was developed under Brehmer's supervision. And now, we have arranged to conduct a series of paintball experiments with Swedish soldiers as participants.

A platoon at the Kvarn basic training camp has been assigned to assist in conducting the experiments. These soldiers are blind to the hypotheses being tested. They will act as observers and data recorders. They will man the stop-watches to measure response times and count the number of times that targets are hit. They will also assist in the running of the experiment.

Three officers have volunteered to play the role of leader. The officers are also blind to the hypotheses being tested.

Five platoons of soldiers will participate in the experiment over the course of four days. These soldiers will not know the officer who acts as their leader in the experiment.

This arrangement should satisfy reviewers who have pointed out that the procedure used previously ran the risk of being influenced by experimenter bias. Previously, the time-keepers were not blind to the hypotheses.

We will use a repeated measures design with three conditions: present leader, remote leader, and remote leader with buzzer. In the present leader condition, the leader is physically present in the assault lane with the soldier. In the remote leader condition, the leader sits in a separate building where he can observe the soldier's progress. In the buzzer condition, the remote leader activates a cutaneous communication device – a buzzer placed between the soldier's shoulder blades – when he issues his orders.

In all three conditions the leader and the soldier communicate using a radio with a voice-activated headset. The leaders uses the radio to issue verbal commands to move and to shoot. The soldier uses the radio to tell the leader when he is out of ammunition and when he has finished refilling.

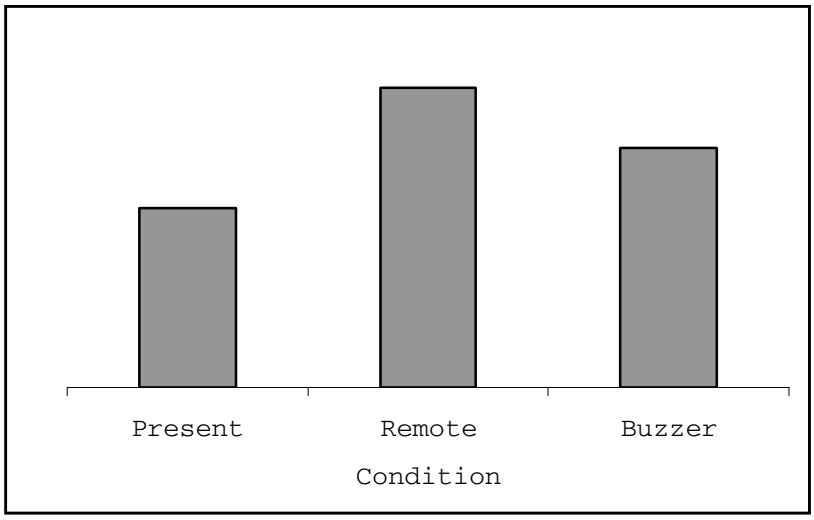
Dependent variables are (a) the soldier's response time to commands to move and to shoot, (2) shooting accuracy, (3) a questionnaire on trust in the leader.

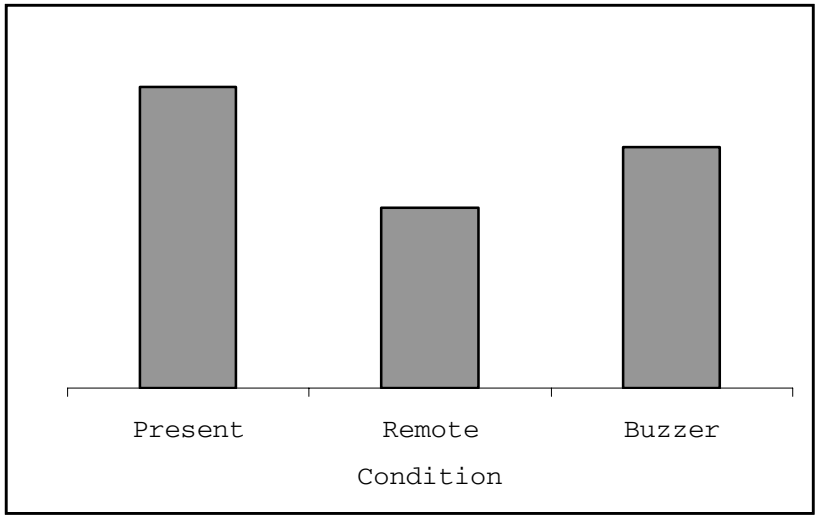
The hypotheses being tested are:

1) Leader presence matters.
   a) Soldiers will be faster to obey verbal commands when the leader is present
   b) Soldiers will shoot more accurately when the leader is present
   c) Soldiers will indicate higher levels of trust in the leader when the leader is present

2) Cutaneous communication fosters trust in the remote leader.
   a) Soldiers will be faster to obey the remote leader's commands in the buzzer condition
   b) Soldiers will shoot more accurately in the buzzer condition than in the remote condition
   c) Soldiers will indicate higher levels of trust in the buzzer condition than in the remote condition

These hypotheses suggest that a plot of the data will take the following form, as thy have in our previous experiments:


Present
Remote
Buzzer
Condition


Present
Remote
Buzzer
Condition

Linköping Institute of Technology Dnr IKP 214/03-23
Final Summary Report
& Progress Report covering the period 1 April 2006 – 31 July 2006
MURI OPUS -- DAAD19-01-1-0621

Principal Investigator: Kip Smith, Ph.D.
Division of Industrial Ergonomics

The focus of research for the MURI OPUS grant at Linköping University and Kansas State University has been the empirical investigation of the effect of trust on the cognitive performance of individuals and small teams working in stressful conditions. There have been two complementary programs of research: (a) trust between team-members and (b) trust in the team's remote leader.

Both programs directly address the Army's interest in the effect of stressors on soldier cognitive performance on the electronic battlefield. The advent of remote command and control and its associated reliance on electronic communications dramatically increase the likelihood that soldiers will find themselves isolated on the battlefield (e.g., Bowden, 1999). This isolation can be a source of benumbing stress (Marshall, 1947). In these conditions, the soldier's only link to his platoon and his leader(s) is their mutual trust. In our experiments we investigate the impact of (a) trust and trustworthiness and (b) leader presence on individual cognitive performance and teamwork.

Our hypothesis in all experiments has been that trust facilitates cognitive performance. Our dependent measures have been (a) success at finding the ill-defined path to a well-defined goal and (b) the speed with which a rifleman responds to his platoon leader's orders to move and to shoot. The first is a measure at the level of the team, the second a measure at the level of the individual soldier.

Every one of our experiments has supported our hypothesis about the beneficial effects of trust on cognitive performance under stress.

Trust between team members

The first program of research tested the hypothesis that revealed trust and trustworthiness between members of small teams (dyads) have a positive influence on subsequent teamwork. The corollary hypothesis is that revealed distrust or betrayal has a negative influence on subsequent teamwork.

Groups of four volunteers were assigned to separate rooms where they were randomly and anonymously paired into two teams of two. The anonymous pairing reduced the likelihood of reputation effects and familiarity.

The experimental procedure had two parts. In the first part of the experiment, each pair was presented with the extensive-form trust game shown in Figure 1. The extensive-form trust game is a standard paradigm in experimental economics (e.g., McCabe, Rassenti, & Smith, 1996; McCabe, Smith, & LePore, 2000). One participant was randomly assigned the role of Player 1, the other Player 2. After receiving instructions about how to play, the pair played the game once. The game has three possible outcomes: distrust, trust and trustworthiness, trust and betrayal. Following the tradition of experimental economics, we paid both participants in cash

immediately upon completion of the game. Our innovation was to ascertain the influence of its three different outcomes on subsequent teamwork and decision making.

In the second part of the experiment, each pair worked together (anonymously and remotely) to play one of two computer games (Hobbits and Orcs or the Boat Game) against a clock. Both games required the team to cooperate under time pressure to achieve a well-defined goal.

This experimental paradigm addresses a fundamental question about teamwork on the electronic battlefield: if platoons of future force warriors are assembled ad-hoc and do not know each other, will they trust each other and will they work effectively to achieve the team's mission?

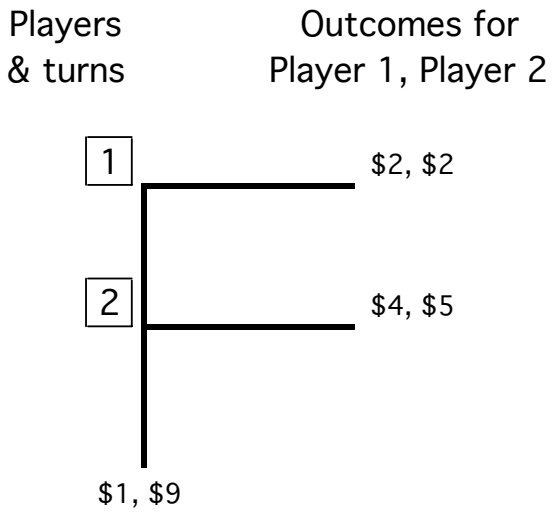


Figure 1. The extensive-form trust game used as the stimulus in the quasi-manipulation of trust. The numbers in the squares represent Player 1 and Player 2. Player 1 has the first move. There are three possible outcomes.
Distrust: A move to the right by Player 1 is a distrusting move that nets both players $2. The distrusted Player 2 does not get to make a move.
Trust and Trustworthiness: A move down by Player 1 is a trusting move that passes play to Player 2. A move right by Player 2 is a trustworthy move that nets Player 1 $4 and Player 2 $5.
Trust and Betrayal: A move down by Player 1 is a trusting move that passes play to Player 2. A move down by Player 2 betrays Player 1's trust. It nets Player 1 $1 and Player 2 $9.

Our experiments revealed that pairs that exhibited trust and trustworthiness in the extensive-form trust game went on to perform better in the games with well-defined goals than teams that experienced either distrust or betrayal. They were reported in the series of quarterly reports, were published in journals and proceedings, and were presented to a wide range of audiences at several professional conferences (Bloomberg, Hawkins, Davidson, Swan, & Smith, 2003; Fosse, & Smith, 2002; Fosse, Murphy, Perry, Ratcliff, & Smith, 2002; Pruitt, N., Bolton, Kearney,

Swan, & Smith, 2003; Ratcliff, & Smith, 2003; Swan, Shane, Ratcliff, Pruitt, Kearney, Freund, Fosse, & Smith 2002; Swan, Younger, Smith, Toll, & Smith, 2003).

This program of research documented that teams of isolated individuals perform better at their shared task if they have had the prior experience of mutual trust. The implications for the Army and the electronic battlefield are direct: the future force warrior needs to know and trust the members of his platoon and know, in turn, that those members trust him. This result argues against the ad-hoc formation of fire-teams. No amount of electronic gadgetry can overcome the insidious effects of distrust (Marshall, 1947).

Trust in the leader and the effect of remote command and control

The second program of research tested the hypothesis that leader presence in the field of fire matters and that trust is the intervening variable between leader presence and soldier performance. The specific hypothesis that was tested was: a rifleman will (a) be quicker to obey verbal orders received over a radio from his platoon leader and (b) trust the leader more if the leader is collocated in the field of fire than if the leader is in a remote location. The corollary hypothesis is that a rifleman's performance and his trust in his platoon leader is degraded under conditions of remote command and control.

Four experiments were conducted, one in the United States and three in Sweden. The experimental setting for all four was the paintball assault lane like that illustrated in Figure 2. The assault lane was delineated by two ranks of alternating, equally-spaced protective barriers. A small cup containing five paintballs was placed behind each barrier. At the far end of the lane stood a protective wall that formed a sniper's hardened position. A group of targets were placed in random order in front of the sniper's position.

Participants played the role of rifleman. In the first three experiments, participants were men and women 18 to 25 years old. In the fourth experiment, participants were three platoons of active duty Swedish Army soldiers. Each rifleman's task was, individually, to follow his leader's orders to move (advance up the file lane), to shoot (to attempt to hit the targets), and to avoid being hit by the sniper's fire. The assault lane paradigm was borrowed from U.S. Army training procedures.

All four experiments used two conditions of leader presence in a repeated measures design to test the effect of leader presence on response times to direct verbal orders. In the collocated condition, the leader was present in the assault lane one barrier behind the rifleman and was, on occasion, hit by the sniper's fire. In the remote condition, the leader stood in a safe location (out of the range of the sniper's fire) where he could see but could not be seen by both the rifleman and the sniper. The leader was unknown to the participants. In both conditions, the same leader issued verbal orders to the rifleman using a radio. The order of conditions was counterbalanced. The leader provided no covering fire and did not model for the rifleman how to advance up the lane or to fire his weapon. The dependent variable was the time it took the rifleman to obey an order from the leader to move ('move times') and to fire ('fire times').

Each experiment progressively tightened controls on potential confounding variables. The same result was observed in all four experiments: riflemen were significantly faster to obey

commands issued by a collocated leader than by (the same but) remote leader. In the fourth experiment, a new dependent variable was added: a self-report questionnaire on the rifleman's trust in the leader.

In Figure 3 we plot the difference between the rifleman's two assessments of trust on the horizontal axis and the difference between his two move times on the vertical axis. Greater trust in the collocated-leader condition generates negative numbers on the horizontal axis. Faster times in the collocated-leader condition generate positive numbers on the vertical axis. Thus, our hypothesis leads us to expect this cross-plot to have a downward sloping trend.

To test that hypothesis, we conducted a simple linear regression analysis. The resulting model is: Difference in time = 0.40 - 0.30 * Difference in trust. A one-sided t-test for the slope parameter (β_1, -0.30) is significant, $t(52) = -2.312$, $\underline{p} < .012$. Analysis of residuals indicates no deviation from normality and satisfactorily constant variance at all levels of the predictor variable (trust). While trust explains only 9% of the variance in the data ($r^2 = .09$), the model supports the hypothesis that there is a significant association between trust and performance, with soldiers trusting less and responding more slowly in the remote-leader condition.

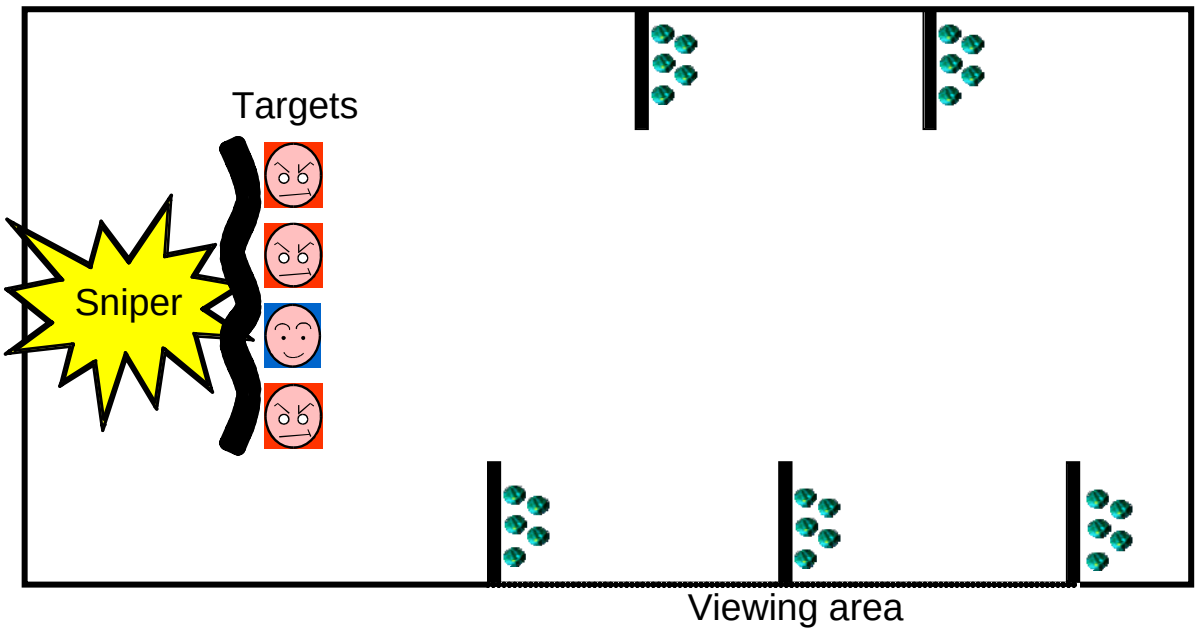


Figure 2. The basic layout of the paintball assault lanes used in the experiments on the effect of leader presence on soldier performance. Two staggered ranks of barriers formed the lane. At the head of the lane was a sniper's hardened position. The barriers provided the soldier protection from the sniper's fire. A cup containing five paintballs was located at each barrier, enabling the soldier to reload. Targets, both friendly (blue) and enemy (red), sat in front of the sniper's bunker.

This program of research revealed that riflemen (a) trust a collocated leader more than a remote leader and (b) obey a collocated leader faster than a remote leader. The correlation between the two suggests, but does not conclusively demonstrate, that trust is the intervening variable that mediates leader presence and soldier performance.

The experiments were reported in the series of quarterly reports, were published in journals and proceedings, and were presented to a wide range of audiences (Liberg & Smith, 2006; Pangburn, Freund, Pangburn, & Smith, 2003; Smith, 2004a, 2004b; Smith, & Liberg, 2005) Additional publications are in preparation.

The implications for the Army and the electronic battlefield are direct: the future force warrior, like all soldiers throughout history (Marshall, 1947), will be ill-served by remote command and control of fire teams.

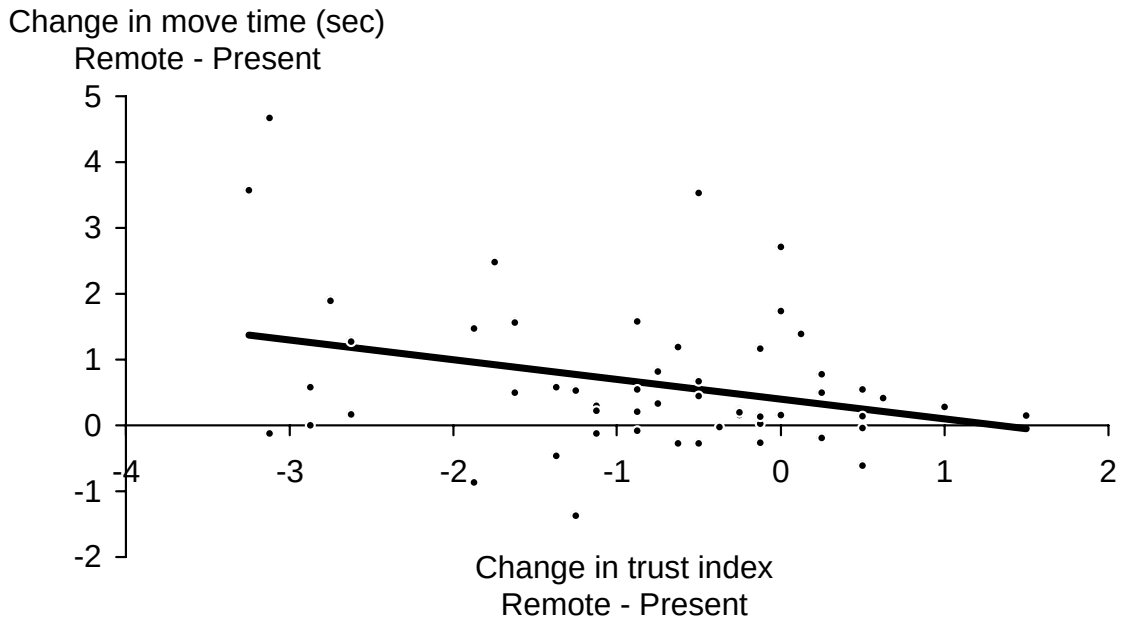


Figure 3. Scatter plot and best-fit linear regression function showing the significant negative association between the difference in experienced trust and the difference in move time.

Second quarter 2006 report

As reported in the first quarter 2006 report, all experimentation was abruptly terminated at the request of the University of Central Florida due to an unfortunate bureaucratic misunderstanding about Linköping University's Federal Wide Assurance Number. The notification that we were to stop conducting experiments for the MURI project had a significant 'trickle-up' effect. It appears that not only Linköping University but the entire Swedish research community had been

operating without such a number for some time. The lawyers went to work and managed to resolve the crisis this month (July 2006). We have been told that Sweden generally and Linköping University specifically now have a new and valid Federal Wide Assurance Number.

Unfortunately, this progress comes too late for the MURI OPUS grant. Funding will lapse at precisely the time when it would have been permissible for experiments to restart.

The University of Central Florida set Linköping University's budget for 2006 at less than 50% of the amount specified in the original grant proposal. Given the limited budget and the ban on data collection, the only possible work was writing. The first paper was attached to the first quarter report. During this quarter it was accepted for presentation at the 50th Annual Meeting of the Human Factors and Ergonomic Society. The second and third papers are in progress. The second is a book chapter reviewing the entire set of paintball experiments. The third is a journal article discussing the capstone experiment conducted with active-duty Swedish Army riflemen. Drafts of these articles will be forwarded to the University of Central Florida before the beginning of the fall 2006 semester.

Cited publications by others

Bowden, M. (1999). Black hawk down. New York: Signet Press.

Marshall, S. L. A. (1947). Men against fire. New York: William Morrow & Co.

McCabe, K., Rassenti, S., & Smith, V. (1996). Game theory and reciprocity in some extensive form experimental games. Proceedings of the National Academy of Science **93**, 13421-13428.

McCabe, K., Smith, V., & LePore, M. (2000). Intentionality detection and "mindreading": Why does game form matter? Proceedings of the National Academy of Science **97**, 4404-4409.

MURI supported publications by Dr. Smith acknowledging MURI support

In preparation

Smith, K. (In preparation). Remote command and control, trust, stress, and soldier performance. In Hancock & Szalma, Performance Under Stress. Ashgate.

Smith, K., Liberg, M., and Pangburn, K. (In preparation). The value of leader presence and trust to men against fire.

Journal Articles

Smith, K. (2004a). The adverse impact of remote command and control under live fire. Human Performance in Extreme Environments.

Presentations and Proceedings

2006

Liberg, M., & Smith, K. (2006). Remote command and control compromises soldiers' trust in their leaders. Accepted for presentation at the Human Factors and Ergonomics Society 50th Annual Meeting. San Francisco, CA. October 16 – 20.

2005

Smith, K., & Liberg, M. (2005). Paintball 2: A study of remote-controlled tactile stimulation for the simulation of leader presence. Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society, 364 - 368.

2004

Smith, K. (2004b). Initial experiments on leader presence and communication mode on combat performance. Proceedings of the Second Conference on Human Performance, Situation Awareness, and Automation Technology 2, 266 - 271.

2003

Bloomberg, S., Hawkins, S., Davidson, R., Swan, B., & Smith, K. (2003). Distributed teams and revealed trust. Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, 678 - 682.

Pangburn, K., Freund, S., Pangburn, H., & Smith, K. (2003). Beyond Milgram: An experimental study of leader presence. Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, 364 - 368.

Pruitt, N., Bolton, J., Kearney, A., Swan, B., & Smith, K. (2003). Hobbits and Orcs: communication, trust, and team problem solving. Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, 686 - 690.

Ratcliff, E., & Smith, K. (2003). Heart rate, emotion, and price formation in experimental auctions. 15th Annual Meeting of the American Psychological Society.

Swan, B., Younger, E., Smith, J., Toll, M., & Smith, K. (2003). Communication, proximity, and trust in distributed command and control. Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, 414 - 418.

2002

Fosse, N., & Smith, K. (2002). The effect of trust on cooperative behavior. 2002 Sigma Xi Student Research Conference. Galveston, TX.

Fosse, N., Murphy, L., Perry, J., Ratcliff, E., & Smith, K. (2002). Preliminary studies of the impact of an economic manipulation of trust on team problem solving. Oklahoma-Kansas Society for Judgment and Decision Making. Oklahoma City, OK.

Swan, B., Shane, A., Ratcliff, E., Pruitt, N., Kearney, A., Freund, S., Fosse, N., & Smith, K. (2002). The impact of an economic manipulation of trust on team problem solving. 2002 Annual Meeting of the Society for Judgment and Decision Making. Kansas City, MO.

Accepted for publication in *Human Factors*

Designing for Flexible Interaction between Humans and Automation: Delegation Interfaces for Supervisory Control

Christopher A. Miller
Smart Information Flow Technologies
Minneapolis, MN U.S.A.

Raja Parasuraman
George Mason University
Fairfax, VA U.S.A.

Objective: We develop a method for enabling human-like, flexible supervisory control via delegation to automation. ***Background***: Real time supervisory relationships with automation are rarely as flexible as human task delegation to other humans. Flexibility in human-adaptable automation can provide important benefits including: improved situation awareness, more accurate automation usage, more balanced mental workload, increased user acceptance, and improved overall performance. ***Method***: We review problems with static and adaptive (as opposed to "adaptable") automation. We contrast these approaches with human-human task delegation, which can mitigate many of the problems. We revise the concept of a "level of automation" as a pattern of task-based roles and authorizations and argue that delegation requires a shared hierarchical task model between supervisor and subordinates, used to delegate tasks at various levels and offer instruction on performing them. A prototype implementation, called Playbook^{TM}, is described. ***Conclusion***: Delegation offers a viable method for flexible, multi-level human-automation interaction to enhance system performance while maintaining user workload at a manageable level. ***Application***: Most applications of adaptive automation (aviation, ATC, robot control, process control, etc.) are potential avenues for the adaptable, delegation approach we advocate. We present an extended example for UAV mission planning.

Address correspondence to: Christopher A. Miller, SIFT Inc., 1272 Raymond Ave., St. Paul, MN 55108; cmiller@sift.info

Short, Running Title: Flexible Delegation for Supervisory Control

Keywords: automation, adaptable automation, adaptive automation, delegation, Playbook^{TM}, supervisory control, unmanned air vehicles, levels of automation

INTRODUCTION

As systems become more complex, there is often a need to control them via automation—that is, using a device, machine, or system that accomplishes, fully or partially, a function which was or could be performed by a human (Parasuraman, Sheridan, & Wickens, 2000). Examples of this trend are rife in domains ranging from aviation to health care (Billings, 1997; Degani, 2004; Parasuraman & Mouloua, 1996; Sheridan, 2002).

Automation provides clear benefits and many complex human-machine systems cannot be operated successfully without it. Billings (1997) documents payoffs of aviation automation in four key areas: safety, reliability, economy, and comfort. But automation also poses novel problems for operators. Poorly designed automation can increase workload and training requirements, decrease situation awareness, and, in extreme circumstances, lead to accidents (e.g., Billings, 1997; Casey, 1993; Degani, 2004; Parasuraman & Riley, 1997; Reason, 1997; Sarter, Woods, & Billings, 1997; Vicente, 2003).

This "double-edged sword" of automation use (Bainbridge, 1983) has motivated repeated questions about what to automate to what level or degree for optimal control, performance, and safety. Technologists tend to push to automate tasks as fully as possible—the so-called "technological imperative" (Sheridan, 1987). Human factors engineers and others concerned with safety and the human role in advanced systems have tended to highlight the risks of increased automation (e.g., Perrow, 1984; Reason, 1997; Woods, 1996) and to argue against the use of higher levels of automation.

An approach to human-automation relationships that retains the benefits of automation while minimizing its costs and hazards is needed. For reasons discussed below, we believe that such an approach requires that neither human nor automation be *exclusively* in charge of most tasks, but rather uses intermediate Levels of Automation (LOAs) and flexibility in the role of automation during system operations, and places control of that flexibility firmly in the human operator's hands. Operators need to be able to delegate tasks to automation, and receive performance feedback, in much the same way that delegation is performed in successful human-human organizations¬: at various levels of detail and granularity, and with various constraints, stipulations, contingencies, and alternatives (see Milewski & Lewis, 1999, for a review).

The use of intermediate LOAs as a way of improving human-automation performance has been proposed before (Endsley & Kiris, 1995; Wickens, Mavor, Parasuraman, & McGee, 1998). However, these previous proposals saw the choice of an intermediate LOA as a *design* option (for an exception, see Moray, Inagaki, & Itoh, 2000). Once a LOA was designed into a system, it could not be changed during operation. In contrast, we propose that the LOA should be adjustable *during* system operations. This is consistent with the adaptive automation concept (Inagaki, 2003; Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992; Rouse, 1988; Scerbo, 1996). However, many users may be unwilling to accept system-driven adaptation, and any such adaptive changes may increase the system's unpredictability to the user (Billings & Woods, 1994). Previous research has demonstrated the effectiveness of adaptive automation, but in most of this work the automation decides how to adapt its own behavior (Banks & Lizza, 1991; Colucci, 1995; Dornheim, 1999; Duley & Parasuraman, 1999; Hancock, Chignell & Lowenthal, 1985; Kaber & Riley, 1999; Miller & Hannen, 1999;; Parasuraman, 1993; Parasuraman, Mouloua & Hilburn, 1999; Maybury & Wahlster, 1998; Prinzel, Freeman, Scerbo, Mikulka & Pope, 2003; Rouse, 1988; Scallen, Hancock & Duley,

1995; Scerbo, 2001). In contrast, we propose that the human should remain in charge, deciding how much automation to use. This approach has been characterized as *adaptable* automation (Opperman, 1994; Scerbo, 2001). We believe that adaptable automation can lead to similar benefits as adaptive automation, while avoiding many of its pitfalls.

In this paper, we propose and provide supporting evidence for a delegation approach to adaptable human-automation interaction. We first examine traditional LOA characterizations and discuss why a delegation approach demands they be extended. We then present one specific implementation of a delegation approach, based on a sports team "playbook" metaphor. (The term Playbook™ as a human-automation interaction application, is trademarked by Smart Information Flow Technologies, LLC.) We describe the major characteristics of this theoretical framework for supervisory control and outline the benefits for human and system performance.

INTERMEDIATE LOAS

Costs of Automation Extremes

The promised, and frequently realized, benefits of automation are generally sufficient to drive developers away from the lowest LOA: purely manual performance. The economic benefits of automation are a strong, but not the only, motivator. Automation also offers substantial benefits in human workload reduction, improved performance and safety when it is properly designed and used (Billings, 1997; Bright, 1955; Lorenz, Nocera, Rottger, & Parasuraman, 2002; Parasuraman & Riley, 1997; Parasuraman et al., 2000; Reason, 1997).

The case against always automating at the highest level technologically feasible has been harder to make. Nevertheless, much research shows disadvantages from reducing the human role in system control and problem solving (Amalberti, 1999; Bainbridge, 1983; Billings, 1997; Lewis, 1998; Parasuraman & Riley, 1997; Parasuraman et al., 2000; Rasmussen, 1986; Sarter et al., 1997; Satchell, 1998; Sheridan, 1992; Wickens et al., 1998; Wiener & Curry, 1980).

High LOAs, particularly of decision-making functions, may reduce the operator's awareness of system and environmental dynamics (Endsley & Kiris, 1995; Kaber, Omal, & Endsley, 1999). Humans tend to be less aware of changes when those changes are under the control of another agent (whether automation or human) than when they make them themselves (Wickens, 1994). Mode errors also illustrate the impact of automation on awareness of system characteristics (Sarter & Woods, 1994, 1995). Another major issue influencing how effectively human operators use automation is trust (Lee & Moray, 1992, 1994). Operators may not use well-designed, reliable automation if they believe it to be untrustworthy, or they may continue to rely on automation even when it malfunctions if they are overconfident in it. Highly automated systems are prey to both sorts of errors (Parasuraman & Riley, 1997).

If higher LOAs can result in complacency and loss of awareness, it is perhaps not surprising that they can also result in skill degradation. The pilots of increasingly automated aircraft feared this effect with regards to psychomotor skills such as aircraft attitude control (Billings, 1997), but it has also been demonstrated for decision making skills (Kaber et al., , 1999).

Automation can also produce workload extremes, both low and high. That high LOAs could leave an operator bored is, perhaps, understandable, but that automation can increase workload, is one of the "ironies of automation" (Bainbridge, 1983) since many automated systems are introduced as workload-saving moves. If automation is implemented in a "clumsy" manner (e.g., if executing an automated function requires extensive data entry when operators are already very busy), workload reduction may not occur where it is most needed (Wiener, 1988). Also, if engagement of automation requires considerable "cognitive overhead" (Kirlik, 1993; Parasuraman & Riley, 1997) to evaluate the benefit of automation versus the cost of performing the task manually--then users may

experience greater overall workload in using the automation.

The most significant effect of using too high a LOA may be degraded overall performance of the human + machine system. In an experiment involving aircraft route planning, Layton, Smith, and McCoy (1994) provided operators with one of three levels of support ranging from "sketching only," where the human sketched a desired route and the system provided feasibility feedback, to "full automation" where a recommended best route was automatically provided. An intermediate level allowed the user to ask for a route with specific characteristics that the system then provided if possible. Operators in the intermediate and high automation conditions were found to explore more routes because the manual sketching condition was too difficult to allow much exploration and, in the full automation condition, users tended to accept the first route suggested without exploring it or its alternatives deeply. Even when they did explore, the system's recommendation tended to narrow and bias their search. Particularly in trials when the automation performed suboptimally (e.g., because it failed to adequately consider uncertainty in weather predictions), humans using the intermediate LOA produced better overall solutions.

Finally, high LOAs frequently lack user acceptance, perhaps particularly from the highly skilled operators of complex, high-criticality systems such as aircraft, military systems, process control, etc. For example, Miller (1999) interviewed several pilots and designers to develop a consensus list of prioritized goals for a "good" cockpit configuration manager for the Rotorcraft Pilot's Associate (Colucci, 1995), an adaptive automation system for managing information displays and cockpit functions to conform to pilot intent. Though the participants were generally enthusiastic about the technology, two of the top three items on the consensus list were "Pilot remains in charge of task allocation" and "Pilot remains in charge of information presented." Similarly, Vicente (1999) and Bisantz, Cohen, Gravelle, and Wilson (1996) cite examples of human interactions with even such low-criticality automation as food planning aids in fast food restaurants, showing that operators can become frustrated when forced to interact with automation that removes their authority to do their jobs in the best way they see fit. Vicente (1999) summarizes findings from Karasek and Theorell (1990) showing that jobs in which human operators have high psychological demands coupled with low decision latitude (the ability to improvise and exploit one's skills for job performance) lead to higher incidences of heart disease, depression, pill consumption, and exhaustion.

More complete reviews of issues in human interaction with automation may be found elsewhere (Degani, 2004; Parasuraman and Riley, 1997; Parasuraman et al., 2000; Sheridan, 2002; Vicente, 1999; Lee & See, 2004).

Tradeoff Space for Effects of LOA

The findings summarized above show that a mixture of human and automation involvement is frequently desirable rather than the extremes of full or no automation. Thus, human + machine systems must be designed for an appropriate relationship, allowing both parties to share responsibility, authority and autonomy in a safe, efficient and reliable fashion. The effects of different human-automation relationships over a task or function can be viewed as a tradeoff space. Figure 1 presents a conceptual view of this space over three relevant parameters:

1. The *competency* of the human-machine system. Competency refers to correct behavior in context. Therefore, a system becomes more competent whenever it provides correct behaviors more frequently or encompasses a greater number of contexts (including by making finer distinctions between contexts).
2. The mental *workload* for the human to interact with the system. Workload refers to the attentional and cognitive "energy" the human must exert to use the system (Moray, 1979; Parasuraman & Hancock, 2001; Wickens & Hollands, 2000). Mental workload for the human is modeled because

it is the major constraint on the other two dimensions.

3. The *unpredictability* of the system to the human operator. Unpredictability refers to the inability of the human to know exactly what the automation will do when. Unpredictability is a consequence of the human not *personally* taking all actions in the system—of not being "in control" directly and immediately. Unpredictability is inversely related to the workload required to maintain awareness of the actions being taken with the system and generally, and with those specific aspects of situation awareness which pertain to the understanding of the automation behaviors and the system functions they control. Doing tasks directly costs more workload, but the payoff is greater awareness of how the task is being done. For example, while driving a car, a driver's awareness of the behavior of the car, the roads, hazards, etc. will likely be very good, though his or her workload will be comparatively higher than if someone else were driving. Good system and interface design, improved feedback to the user about automation (Nikolic & Sarter, 2000; Parasuraman, 2000), increased reporting requirements and good hiring and training practices can all serve to reduce unpredictability. However, in general, *any* form of task delegation—whether to automation or other humans—must necessarily result in added unpredictability if it offloads tasks.

The triangle in Figure 1 illustrates the relationship, or "tradeoff space," between these three dimensions. Other dimensions are certainly relevant to achieving competency, including: operator skill, competencies and training, understanding of the domain, fatigue and environmental stressors, etc. We have emphasized these three dimensions as both relevant to supervisory control interactions and amenable to modification during use or interaction.

Performance to a given level of competency can only be achieved through some mix of workload and unpredictability (Milewski & Lewis, 1999). A user's workload can be reduced by allocating some functions to automation, but only at the expense of increased unpredictability (at least with regards to those functions); conversely, reducing unpredictability by having the user perform functions increases workload. It is sometimes possible to reduce both workload and unpredictability for a given level of competency through better design—corresponding to shortening the height of the triangle.

In this tradeoff model, any increase in human

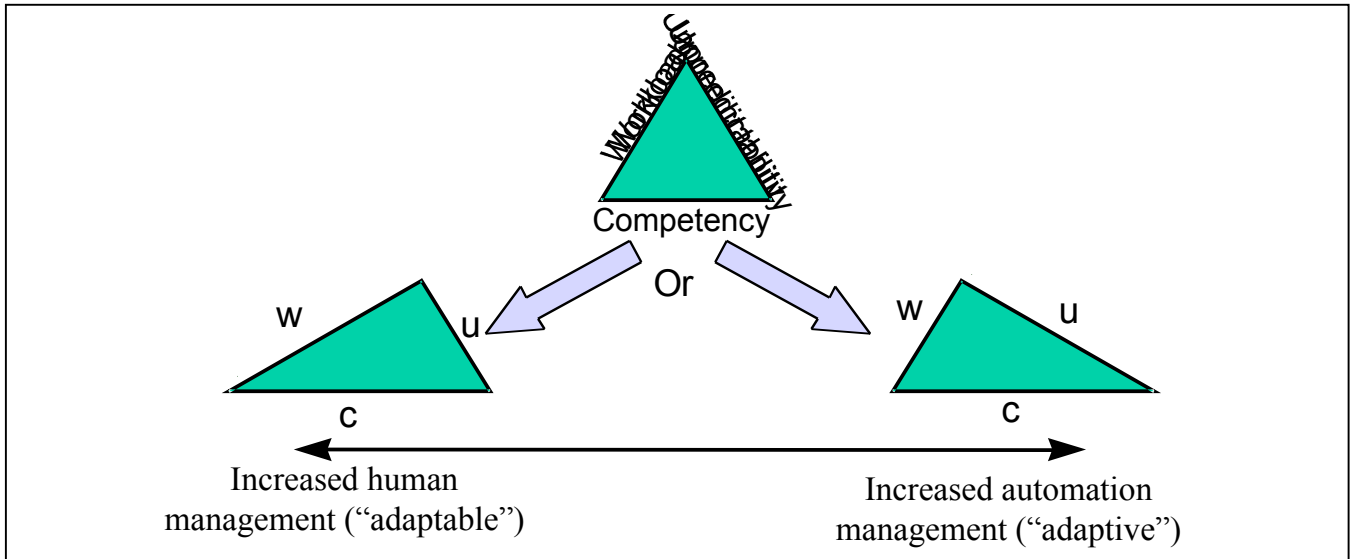


Figure 1. The spectrum of tradeoff relationships between system competency, human workload and unpredictability.

+ machine system competency must affect the human in that either (1) the added functionality must be controlled by the human, resulting in workload increases, or (2) it must be managed by automation, resulting in unpredictability increases. Opperman (1994) identified these alternatives as "adaptive" and "adaptable" approaches to automation usage respectively (see also Scerbo, 2001) depending on whether the automation or the human determines and executes the necessary adaptations respectively.

Another implication of the model is that these approaches are the endpoints of a spectrum with many alternatives in between, each representing a different tradeoff between workload, unpredictability and competency and each a different mix of human and automation roles and responsibilities. The range of alternatives may be constrained by automation and/or human capabilities, but within the feasible range, an alternative must be selected. This selection *is* the process of choosing an automation "level" or relationship—that is, of selecting who will be responsible for which tasks and what the performance constraints, authorizations and reporting requirements will be (Parasuraman et al., 2000). When this division of labor is done by a designer prior to operation, it is part of the design for that system. When it is done by a supervisor for a human team (either prior to an operation or dynamically during it), the process may be called "delegation" or, more generally, "tasking" and task management.

Flexible LOA

Traditionally, the chief difference between task delegation performed by a supervisor and the design of an automated system has been that the supervisor had much more flexibility in what, when and how to delegate and better awareness of the task performance conditions, while the designer had to fix a relationship at design time for static use in all contexts for the to-be-built system. More recently, adaptive automation (Banks & Lizza, 1991; Miller & Hannen, 1999;; Parasuraman, Mouloua, & Molloy, 1996; Rouse, 1988; Scallen et al., 1995; Scerbo, 1996) has been developed which chooses a LOA for tasks based on contextual criteria such as location, workload, experience, physiological state, etc. In such systems, the division of labor between human and machine is not fixed, but varies under the control of the automation.

The adaptive automation concept was proposed over 25 years ago (Rouse, 1976). Only recently, however, have technologies matured to enable empirical evidence of its effectiveness to be provided. Studies have shown that adaptive systems can regulate operator workload and enhance performance, while preserving the benefits of automation (Hilburn, Jorna, Byrne, & Parasuraman, 1997; Kaber & Endsley, 2004; Moray et al., 2000; Prinzel et al., 2003). The performance costs of certain forms of automation described previously—reduced situation awareness, complacency, skill degradation, etc.—may also be mitigated (Kaber & Riley, 1999; Parasuraman, 1993; Parasuraman et al, 1996; Scallen et al., 1995), although if adaptive systems are implemented "clumsily," performance can be even worse than with fully manual systems (Parasuraman et al., 1999). For a recent review of adaptive automation research, see Inagaki (2003). Adaptive aiding systems have been successfully flight tested in the Rotorcraft Pilot's Associate (Dornheim, 1999).

Adaptive automation is truly "adaptive" in Opperman's (1994) sense since it chooses the LOA to apply. By contrast, a supervisor's task delegation within a team is more nearly "adaptable" (Cooke, Salas, Cannon-Bowers, & Stout, 2000), since the supervisor can choose which tasks to give to a subordinate, how much to dictate about how to perform subtasks, how much attention to devote to monitoring, approving, reviewing and correcting task performance, etc.

This ability to delegate tasks to subordinates is clearly a powerful form of organizing and performing work. It is also deeply familiar to anyone who has worked in a team, especially in a supervisory setting. As described below, we have sought to provide delegation-like interactions for adaptable control over flexible automation. Before presenting that work,

however, we must first define a "level of automation" and then argue for an extension to traditional definitions to support delegation-based approaches to supervisory control.

CHARACTERIZING LOAS

To discuss alternative forms of automation, a scheme for characterizing, roles and responsibilities is helpful. Such characterizations have traditionally been made in terms of LOAs: defining a spectrum of possible relationships ranging from full human to full automation control.

Prior Taxonomies for LOAs

Bright (1955), perhaps the first to propose such a spectrum of LOA, viewed automation as evolving through 17 levels of competency and noted that intermediate stages might well demand greater human workload, skill and training requirements than lower levels. Sheridan's (1987; see also Sheridan & Verplank, 1978) seminal definition of the *supervisory control* included the provision that control automation allow the human to behave as if interacting with an intelligent, human subordinate—in Sheridan's words, "…setting initial conditions for, intermittently adjusting and receiving information from a computer that itself closes a control loop (i.e. interconnects) through external sensors, effectors, and the task environment." (Sheridan, 1987, p. 1244). Sheridan also described a 10-point spectrum of automation levels whose endpoints are full control autonomy for the human (essentially no role for automation) and vice versa (Sheridan & Verplank, 1978). The intermediate levels in this spectrum, then, represent alternative forms of supervisory control interactions. A version of this list is shown in Table 1.

Table 1. Levels of Automation (after Sheridan and Verplank, 1978).

1. Human does it all.
2. Computer offers alternatives
3. Computer narrows alternatives down to a few
4. Computer suggests a recommended alternative
5. Computer executes alternative if human approves
6. Computer executes alternative; human can veto
7. Computer executes alternative and informs human
8. Computer executes selected alternative and informs human only if asked
9. Computer executes selected alternative and informs human only if it decides to
10. Computer acts entirely autonomously

A problem with uni-dimensional models of human-automation relationships is that they are ambiguous about what the application domain of the relationship. Parasuraman et al. (2000) noted that Sheridan's levels referred mainly to automation that makes decisions, offers suggestions and/or executes actions. There are, however, other jobs automation can do: for example, sensing and analyzing information to detect situations of interest, without necessarily offering any advice on what to do with the information. Parasuraman et al. (2000) applied a simple, stage model of human information processing to arrive at four functions that must be accomplished to perform most tasks: (1) information acquisition, (2) information analysis, (3) decision and action selection, and (4) action implementation.

Since these functions can be performed by either human or automation in various mixes, in effect Parasuraman et al., (2000) added a second dimension to Sheridan's spectrum—that of the function or task the relationship is defined over. Most human + automation systems can be characterized by a mix of LOAs across these four functions, as in Figure 2. One system (A) might be highly autonomous in information acquisition, but comparatively low on the other functions, while a second (B) might offer a high LOA across all four functions. For example, some flight booking websites currently provide a high LOA for information acquisition—searching across multiple airlines and other websites—while a few also perform some

information analysis (sorting flights by their cost or proximity to desired departure times). None currently provides specific recommendations, much less actually executes (autonomously) the ticket purchase, but these capabilities are easily envisioned and technologically feasible.

Extending LOAs

An implication of the Parasuraman et al. (2000) levels x functions model is that a parent task can be decomposed and that a single automation level need not be applied homogenously across the subtasks. However, in this model a parent task is decomposed into abstract subfunctions based on information processing stages, whereas other decomposition methods might arguably provide more insight into how a task may be performed. Alternatives include:

- Finer-grained functional decompositions such as Operator Functional Modeling (Mitchell, 1987), Goals-Operations-Methods-Scripts (GOMS) models (Card, Moran, & Newell, 1983), and Plan-Goal Graphs (Geddes, 1989) that stress the sub-functions causally required to achieve a parent, and/or
- Sequential process models such as PERT or CPM charts (e.g., Martinich, 1996) and Petri nets (Murata, 1989) that stress the temporal ordering and duration of sub-functions.

Such decompositions are inherently hierarchical and may proceed through any number of levels to some primitive, "stopping" level (Diaper, 1989) that may be imposed by biology, physics or, more commonly, the purpose of the decomposition.

Thus, while the two-dimensional LOA model offered by Parasuraman et al. (2000) represents a major advance over earlier uni-dimensional models, it arguably does not go far enough. The subdivision of a parent task into four information processing phases represents only a single level of decomposition into abstract task categories. In practice, tasks are accomplished by hierarchically decomposable sequences of *specific* activities—the parent task's subtasks.

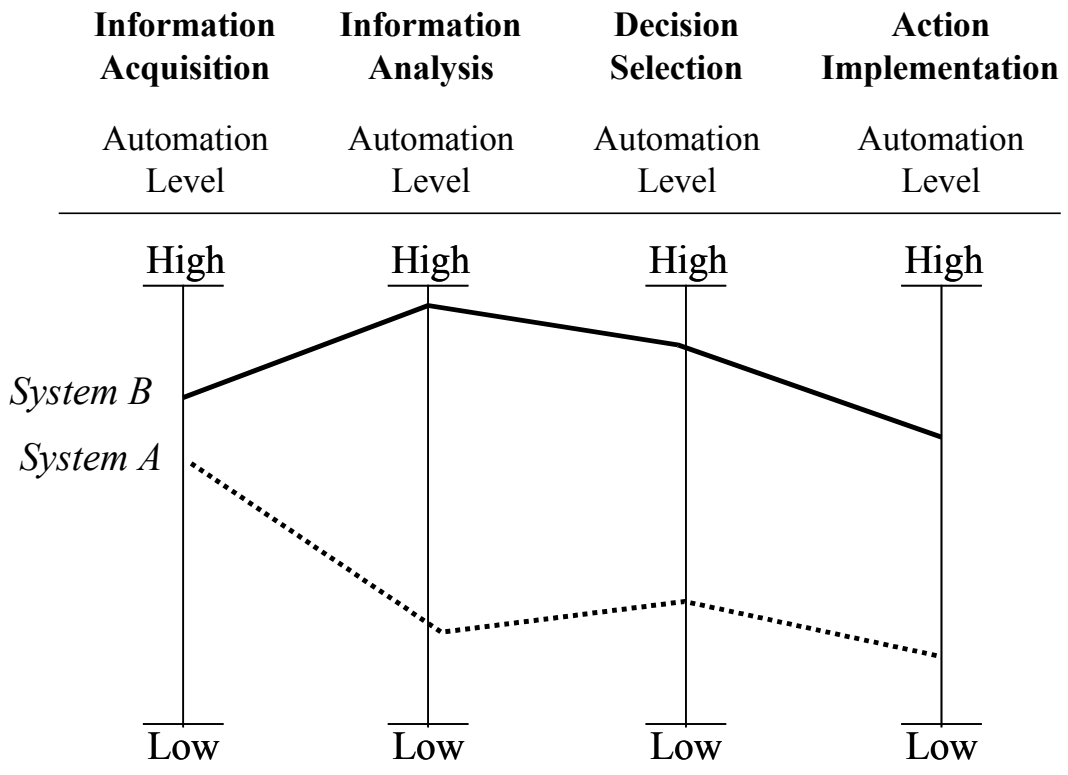


Figure 2. Levels of automation by information processing phase for two systems (from Parasuraman et al., 2000.)

Automation may be applied differently to each and every subtask that comprises the parent task. Thus, the profile of automation levels sketched in Figure 2 could stretch instead over as many subtasks and levels as we want or need to divide a parent task into.

In fact, the relationship between automation "level" and task decomposition is more complex than this. As is well understood (Billings, 1997; Parasuraman et al., 2000; Woods, 1996), automation does not merely shift responsibility for tasks but can change their nature as well. In a task decomposition, this means that some sub-tasks may be eliminated while others are added. This implies that there will generally be multiple alternate decompositions depending on, among other things, what LOA is used. Each alternative constitutes a different combination of human and automation subtasks and, thus, a different method of accomplishing the parent task.

When one identifies a LOA for a complex system using Bright's or Sheridan's dimensions, one is in principle identifying something like an average or modal level over the subtasks the human + automation system accomplishes. Similarly, when one uses a levels x stages model such as in Parasuraman et al. (2000), one is performing an abstract and coarse-grained decomposition of the parent task into subtasks clustered by information processing stage. Assigning levels by subtask stages offers more sensitivity than assigning them only to the parent task, but it is still an abstraction. In practice, one could identify the *specific* subtasks to be performed and represent an automation level for each of them.

Why would one want to? More and finer-grained sub-tasks are not necessarily better and, in fact, Parasuraman et al. (2000) explicitly state that they chose a four-stage model to simplify design considerations. Precision may be inherently desirable for some purposes (such as training and detailed design), but our interest is in supporting flexible task delegation. As we saw above, for any intermediate LOA for a task, there are roles for both humans and automation in its sub-tasks. Yet, someone must coordinate those roles. Insofar as human supervisors are required to manage, or at least be aware of, that division of labor, they must understand the decomposition of the task and of the allocation and coordination requirements among its subtasks. Supervisory control is a process of task delegation and delegation requires task decomposition.

Task decomposition seems to reflect the way humans delegate responsibilities to subordinates, and reason about task performance when receiving feedback (Klein, 1998). Hierarchically decomposed task models, which embody both functional and sequential relationships among tasks, may in fact be an accurate representation of the mental model (Gentner & Stevens, 1986) that supervisors and workers in supervisory control domains use and share (Stout, Cannon-Bowers, Salas, & Milanovich. 1999), but we are not yet making so strong a claim. Instead, we claim merely that such models form a useful structure for discussing work in a domain and, therefore, a useful vocabulary for issuing supervisory control instructions, reasoning about work to be performed and organizing reports on the progress of such work.

In short, delegation is a process of assigning specific roles and responsibilities for the subtasks of a parent task for which the delegating agent retains authority (and responsibility). Furthermore, communication about intent to subordinates is frequently in terms of specific goals, methods and/or constraints on how, when and with what resources the subtasks should be accomplished (Jones & Jasek, 1997; Klein 1998; Shattuck, 1995). For these reasons, it is essential that those subtasks be modeled explicitly if delegation interactions are to occur.

Using Extended LOA in Design

The realization that a LOA is a defined combination of roles and responsibilities between human and automation for a task to be performed, and that that task can generally be decomposed into subtasks with their own LOAs, opens the doors for more explicit communication between human and automation concerning the tasks to be accomplished. We

are striving to push this communication into the run-time environment—more closely emulating delegation in human-human work relationships. Our method allows the operator to smoothly adjust the "amount" or "level" of automation used by permitting, but not requiring, instructions to automation about what specific behaviors it should perform and which behaviors should be left to the human. Supervisors could be expected to make such decisions depending on such variables as time available, workload, decision criticality, trust, etc.

While this does not eliminate the tradeoff between workload and unpredictability presented in Figure 1, delegation mitigates it by allowing operators to choose a point on the spectrum for their interaction with automation. In other words, the decision about how much and what kind of responsibilities to delegate must still be made, but instead of a designer making it at design time (as in traditional automation design), our approach places it in the hands of the operators at execution time—enabling them to behave more like a supervisor in human-human interactions. This strategy follows the advice of both Rasmussen, Pejtersen and Goodstein (1987) and Vicente (1999) to allow the operator to "finish the design" opportunistically in the context of use.

It is critical that such an approach avoid two potential problems. First, it must make the task of instructing automation to behave as desired achievable without excessive workload. Second, it must ensure safe and effective overall behavior. We have created a design metaphor and system architecture that address these concerns. We call the general class of systems that can take instruction at various, flexible levels *tasking* or *delegation interfaces*, because they allow posing a task to automation as one might delegate to a knowledgeable subordinate. Examples we would label tasking interfaces can be found in Jones and Jasek (1997) and Pande and Hayes (2002). Our particular approach we call a *Playbook*—because it uses the metaphor of a sports team's book of approved plays. To date, we have constructed four Playbook prototypes in the domains of pre-mission planning for UAVs, ground mobile robot control, robot-based "capture the flag" game and dynamic tasking of heterogeneous UAVs by dismounted soldiers. Each is described and illustrated in Parasuraman and Miller (in press).

Below, we first describe the Playbook concept as an example of a delegation interface, and provide a specific, detailed example of a prototype Playbook proof-of-concept, ground-based mission planning tool for commanding Unmanned Air Vehicles (UAVs). Finally, we discuss future extensions, open questions and anticipated payoffs of delegation interfaces in more depth.

A PLAYBOOK APPROACH TO TASKING INTERFACES

One aspect of Sheridan's (1987) formulation of supervisory control is the need for a human supervisor to "teach" or program automation based on the requirements of the task and context. Tasking interfaces facilitate instructing automation in specific tasks and methods by creating a shared knowledge structure of tasks and their relationships that serves as the vocabulary for tasking interactions. This same knowledge can be used to improve the safety and efficacy of plans developed by allowing automation to review and critique human planning. Finally, the Playbook approach streamlines delegation by offering a compiled set of plans, or "plays," with short, easily-commanded labels that can be further modified as needed.

We see three primary challenges involved in the construction of any tasking interface:

1. A shared vocabulary must be developed, via which the operator can flexibly pose tasks to the automation and automation can report how it will perform them.
2. Sufficient knowledge must be built into the automation to enable it to make intelligent choices within the instructions provided by the user. To the degree that insufficient intelligence (including awareness) is present in the subordinate system, tasking may still be possible, but

the results will be unsafe, untrustworthy and generally unacceptable.

3. One or more interfaces must be developed to permit inspection and manipulation of the vocabulary to pose and review tasks rapidly and easily.

Figure 3 presents our general architecture for tasking interfaces. While we are still developing applications that achieve the vision below, we have constructed various partial prototypes—all of which obey the basic structure illustrated. One of these prototypes will be described in detail after a general description of the architecture and function of each component in Figure 3.

The three primary components in Figure 3 each address one of the challenges described above. A User Interface (UI) and an Analysis and Planning Component (APC) communicate with each other and with the operator via a Shared Task Model. Since the Shared Task Model is the "language" in which the user and the Playbook's reasoning components interact, it is drawn as encompassing both the UI and APC. The operator communicates tasking instructions as desired goals, tasks, partial plans or constraints, via the UI, using the task structures of the shared task model. The APC is an automated planning system that understands these instructions and (a) evaluates them for feasibility and/or (b) expands them to produce fully executable plans. The APC may draw on special purpose planning tools to perform these functions, though these are not, strictly speaking, a part of the Playbooks we have developed. Outside of the tasking interface, but essential to its use, are two additional components for making momentary adjustments during execution: an Event Handling component for real-time modifications to plans, and vehicle control algorithms themselves. Since the UI, APC, and the Shared Task Model are unique to tasking interfaces, they will be described in detail below.

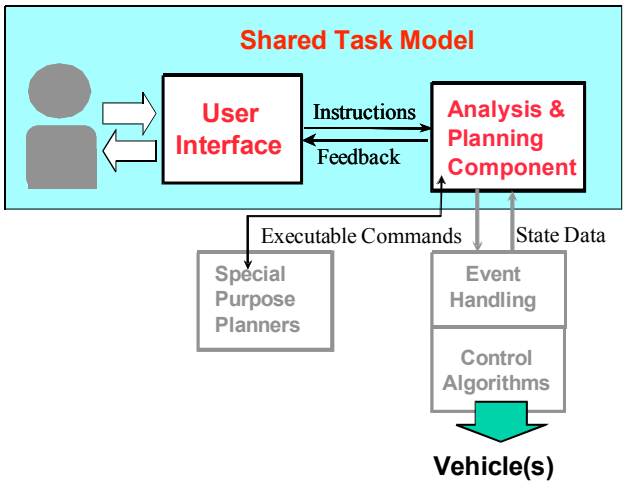


Figure 3. General Tasking Interface Architecture.

Shared Task Model

A critical enabling technology for tasking interfaces is the ability to represent and reason about the goals and plans operators use to talk about work to be performed. By explicitly representing these entities in a format that is familiar to a human yet interpretable by a planning system, we gain a level of coordination beyond that previously possible.

The concept of modeling user tasks is well established in human factors engineering (Diaper, 1989; Kirwan & Ainsworth, 1992). Task modeling techniques are typically comparatively informal (e.g., paper and pencil, or spreadsheet based) means of capturing the steps that knowledgeable users go through to accomplish goals in known circumstances. In essence, they decompose an identified method of goal-accomplishment into a series of partially-ordered steps along with a procedure or script for when and how to employ the steps (Shepherd, 1989). The process can be repeated iteratively, producing a means-ends hierarchy for goal accomplishment.

To be used in a tasking interface, a task model must meet several requirements. First, it must be organized via functional decomposition—each task or goal must be decomposable into alternate methods of achievement, down to some layer of primitive actions which are executable by the event handling and control algorithms of the to-be-controlled system. While this form of decomposition is common in task analysis, it is not universal. Some approaches to using task analyses (especially those for instructional purposes) decompose a single method into

progressively smaller steps without representing alternative methods of achieving the parent goal. For tasking interactions, if there are no alternative performance methods then a flexible tasking interface is not necessary.

Second, the decomposition should be specific, representing executable methods of achieving goals, rather than abstract categories such as information processing stages (Parasuraman et al., 2000). A "task," therefore, represents an encapsulated set of behaviors that is a known method of accomplishing a domain goal. Tasks in the hierarchy therefore provide a vocabulary of goals with associated partial plans and leave degrees of freedom (i.e., which subtask methods to use) for the operator during execution. Thus, this approach partially overcomes criticism (Suchman, 1987; Vicente, 1999) against using static, prescriptive, procedures or scripts in task-based instruction as brittle and impossible to exhaustively specify. While our task hierarchies are, indeed, prescriptive, this is precisely the nature of delegation interactions: to prescribe a goal and a method of its performance. But the method prescribed is never exhaustively stipulated, since some authority must be left to the performing agent if true delegation (and consequent workload reduction) is to occur. In our delegation approach, the task model ceases to be a static artifact around which a system is designed, and instead becomes the language which a human supervisor uses to communicate appropriate performance methods dynamically in the context of use.

Third, tasks must be drawn from the way operators think about their domain and, ideally, should make use of the same labeling conventions operators are familiar with. Task labels and definitions may be drawn from training materials, operator interviews, etc. Finally, the task model must be understandable by the automation using it. Like a football team's playbook, each player must know what to do when a given play is called. The whole task model need not be shared by all members of the team, though both understanding and flexibility will likely be facilitated by greater overlap and sharing between portions of the model, but each must understand his or her portion and the portions must be "linked" in that those pieces which a subordinate understands and will perform when the parent play is called are, in fact, what the superior expects and intends. Coordination is, thus, an emergent function of sharing the same playbook and complicated behaviors can be activated rapidly by use of their labels.

Our task modeling representation is based on a PERT chart-like formalism developed initially by McDonnell-Douglas for use in their Pilot's Associate program (Keller & Stanley, 1992). Using a representation that explicitly models the distinction between goals and plans, such as Geddes' (1989) Plan-Goal Graph, would enhance this approach. The plan-goal distinction is a natural one for users to make in communicating intent (Klein, 1998; Shattuck, 1995).

The operator who "calls plays" must interact directly with the task model, activating and combining tasks at various levels of decomposition. This capability is provided via the user interface described next. We also provide a planning system that can understand the operator's commands and either evaluate them for performability or develop an executable plan that obeys, yet fleshes them out. This Analysis and Planning Component is described subsequently.

User Interface (UI)

Operators must interact with the task model, both to understand possible actions and to declare those tasks they wish the system to pursue. Just as a coach can activate a complex behavior via a simple play name or can spend additional time combining or modifying play elements to create new plays, so operators should be able to tune automation actions via the UI.

Some requirements for the UI include: (1) the set of "plays" (e.g., maneuvers, procedures, etc.) represented must be those any well-trained operator should know, (2) the general play "templates" can be composed and instantiated to create mission-specific plans, (3) the operator may select plays at various hierarchical levels,

leaving the lower levels to be composed by the APC, and (4) operators may either require or prohibit the use of specific plays or resources.

Using the shared task model as an infrastructure, many interfaces are possible—each customized for their usage context. Some useful Playbook UI components are described below. One interface, designed for ground-based UAV mission planning, will be described in depth.

Playbook UIs must include some ability to "call" pre-defined plays from a play library, usually at various hierarchical levels, and thereby command their performance by automation. Play calling may be more or less elaborate according to the demands of the domain. An onboard UI for fighter combat might enable literally "calling" (via a speech interface), simple play-based commands, though most other applications will benefit from more detailed communication. This can be provided minimally by allowing the operator to instantiate the parameters of a play—thereby providing constraints or stipulations on resource usage, initiation time, etc. More elaborately, plays/tasks may need to be composed into longer sequences (e.g., missions). A "play composition" workspace and tool separate from a "play command" tool will help in these cases. Many domains will require creating new plays, either from scratch or by storing the results of earlier play composition. A different tool, or mode, should support this type of interaction. Most domains will need to visualize (and intervene in) the execution of commanded plays. Normal automation interfaces may provide these in a raw form, but referencing performance in terms of the intended plays should improve user understanding. The UI must also support interaction with the APC via issuing partial tasking instructions, receiving critiques, and previewing, accepting and/or modifying APC-generated plans. Finally, unlike the example below, if the Playbook is to be used to monitor and retask automation during execution, then access to performance information, probably referenced against the initial, authorized plan, and ongoing access to fine-grained play commanding capabilities will be necessary.

Interacting directly with an explicit task model meets most of these requirements, but we have found that it helps to make the UI multi-modal. Visualization of the task model shows causal and sequential relationships, but it does not do a good job of conveying the particular assets involved in each task, temporal duration of events, geographical location and progression of events, etc. As we will illustrate in the example below, access to auxiliary tools (such as Gantt charts and interactive map displays) can meet these needs and enhance overall usability.

Analysis and Planning Component (APC)

The APC (Goldman, Haigh, Musliner & Pelican, 2000) operates over full or partial plans provided via the UI to: (1) analyze the operator's plan for feasibility and goal achievement and (2) automatically generate candidate plan completions in keeping with the requirements and prohibitions the human imposes. The APC can critique (Guerlain, 1995) the operator-specified plan for feasibility and constraint violations and it can weed out candidate sub-plays that have been made infeasible by earlier decisions. Finally, the APC can complete (that is, develop to an executable level) a partial plan from whatever level the human chooses. The APC will either incorporate and obey human-specified portions of the plan, or report that no plan can be completed within the constraints.

The APC uses a hierarchical task network planner (Erol, Hendler & Nau, 1994) in conjunction with constraint propagation techniques (Hentenryck, 1989; Jaffar & Michaylov, 1987) to perform the functions described above. Many domain relevant task structures are known before planning begins (e.g., the general steps in a UAV mission). Rather than rediscover this outline from scratch (as in more traditional Artificial Intelligence planning approaches), the APC uses task knowledge from the Shared Task Model. This both ensures that the APC's concepts of

available plays mirrors that of the users and makes its job of creating plans much easier. In turn, the APC must manage the resources, deadlines, etc., checking for feasibility and conflicts. The APC represents these limited quantities as constraints and "feasibility" is defined as the projected plan's ability to achieve the declared goal state (i.e., accomplish the top level task) within resource limitations. As a plan is built, constraints propagate both up, from sub-tasks to tasks, and down, from tasks to their expansions. This constraint propagation process enables the planner to identify flaws in the plan early.

When asked to complete a plan, the APC fleshes out non-primitive tasks by asserting one or more subtask methods that can fill the parent goal. The APC selects its best completion according to resource usage or other optimization criteria. The UI then displays the planning decisions to the user, who can retract choices, or make better-informed decisions. When asked to critique a user-defined plan, the APC again expands subtask methods to an executable level and reports whenever it discovers constraint or goal violations or the lack of any feasible, executable plan. Finally, the APC can also aid user decisions by having feasible plays at the next decomposition level be presented and infeasible ones eliminated or "grayed out," if desired.

In combination, feasibility checking and plan expansion make it possible, but not required, for the APC to generate effective plans with minimal user involvement. Continual feasibility analysis minimizes effort expended on dead-ends while encouraging the user to specify mission critical aspects early. Once these are stipulated, development can (but does not have to) be left entirely to the APC with the assurance that it will produce a plan that is both feasible within its constraint knowledge and in keeping with the operator's stipulations. If time permits (or lack of trust demands), the user may provide increasingly detailed instructions down to the lowest level primitive operators supported.

In fulfilling its role as a plan generator and/or plan checker, the APC operates as a type of decision aiding or decision support system. Unlike traditional decision support systems, however, it operates at a flexible LOA, allowing operators to provide whatever level of tasking they prefer. The shared task model, and the UI which accesses it, operates as a form of programming language which the APC can understand. In this sense, it is similar to the tasking instructions that supervisors can give to subordinates—"programming" them to perform tasks according to the supervisor's intent, but allowing some (flexible) degree of autonomy to accomplish that intent in the prevailing conditions at execution time. As such, the Playbook and its APC are not designed to always reduce workload. The act of providing instructions is potentially workload intensive and, as noted above, Playbook is designed to provide operators the flexibility to decide whether to incur additional workload for reduced uncertainty or not.

Finally, although our early versions of Playbook, including that described below, operated primarily in pre-mission (or "batch") planning, the Playbook concept encompasses the execution and monitoring of plans as well. Recent work (cf. Goldman, Miller, Funk, Wu, & Meisner, 2005; Miller, Goldman, Funk, Wu, & Pate, 2004) has begun to incorporate plan execution capabilities in Playbook. The delegation concept still applies even during execution—the time span between issuing an instruction and the subordinates' execution of it simply shrinks. Stability of control becomes an issue and may limit control authority. Current Playbook implementations have limited capability to autonomously retask assets within the commander's declared intent—for example, adjusting speed or route to maintain critical times on target. Real time execution, monitoring and control via Playbook raises many interesting issues which are, unfortunately, beyond the scope of this paper and which we are only now beginning to tackle.

Auxiliary Components for a Tasking Interface

The Shared Task Model, UI and APC are the not sufficient for most applications. As

illustrated in Figure 3, a Playbook must be integrated with a reactive planning capability for event handling and with low level, real-time control algorithms if it is to enact plays on real world equipment. In practice, this means that tasking interfaces leave some portion of the decisions about how to enact a plan to an intelligent agent at run time. This problem, the gap between analysis/planning and execution, is a familiar one in control and robotics. Many of our implementations have used a reactive planning architecture called CIRCA (Musliner, Durfee & Shin, 1993), which integrates projective and reactive planning with control actuation. CIRCA bridges the analysis/execution gap by synthesizing logically correct reaction programs to achieve high-level goals. However, little thought had been given in CIRCA's development to how an operator would provide those goals. Our approach fills that gap.

At the bottom layer of the architecture (cf. Figure 3), advanced control algorithms provide for continuous vehicle control in a robotics or UAV domain. The plan created jointly by human and APC, along with reactive adaptations provided by the event handling component, gives these control functions the instructions they need.

USAGE SCENARIO

The Playbook approach described has been partially implemented in multiple proof-of-concept demonstration systems (Parasuraman & Miller, in press)--including ones for Tactical Mobile Robots and Unmanned Air Vehicles (UAVs). None of these fully achieves the Playbook vision, but one representative UAV prototype will be described briefly here.

Traditional UAV interfaces either require operators to control the aircraft remotely via a dedicated cockpit mockup, or rely on very high-level behaviors (e.g., Close Air Patrol circuits or waypoint-designated routes) that can be easily but inflexibly commanded. Playbook provides flexibility between these points. Here, we concentrate on a ground-based tasking interface to be used for a priori mission planning. We intend this prototype, however, as an illustration of the general concept and, with appropriate interface modifications, it will be suited to dynamic, real-time and even in-flight tasking as well. The prototype is illustrated in Figure 4.

Figure 4 shows the five primary regions of this Playbook UI. The upper half of the screen is a Mission Composition Space that shows the plan composed thus far. The lower left corner of the interface is an Available Resource Space, currently presenting the set of aircraft available for use. The lower right corner contains an interactive Terrain Map of the area of interest, used to facilitate interactions with significant geographic content. The space between these lower windows (empty at startup) is a Resource in Use Space—once resources (e.g., UAVs, munitions, etc.) are selected, they will be moved to this workspace, where they can be interacted with in more detail. Finally, the lower set of control buttons is always present for interaction. This includes options such as "Finish Plan" for handing the partial plan off to the APC for completion and "Show Schedule" for obtaining a Gantt chart timeline of the activities planned, etc.

At startup, the Mission Composition Space presents the three top-level plays (or mission types) the system currently knows about: Interdiction, Airfield Denial, and Suppress Enemy Air Defenses (SEAD). The mission leader interacts with the Playbook to first, declare that the overall mission play for the day was, say, "Airfield Denial." In principle, the user could define a new top-level play either by reference to existing play structures or completely from scratch, but this capability has not been designed or implemented as yet.

Clicking on "Airfield Denial" produces a menu with options for the user to tell the APC to "Plan this Task" (i.e., develop a plan to accomplish it) or to "Choose airfield denial" as a task that the operator will flesh out further. The pop-up menu also contains a context-sensitive list of optional subtasks that the operator can choose to include under this task. This list is generated by the APC with reference to the structures in the play library, filtered for current feasibility.

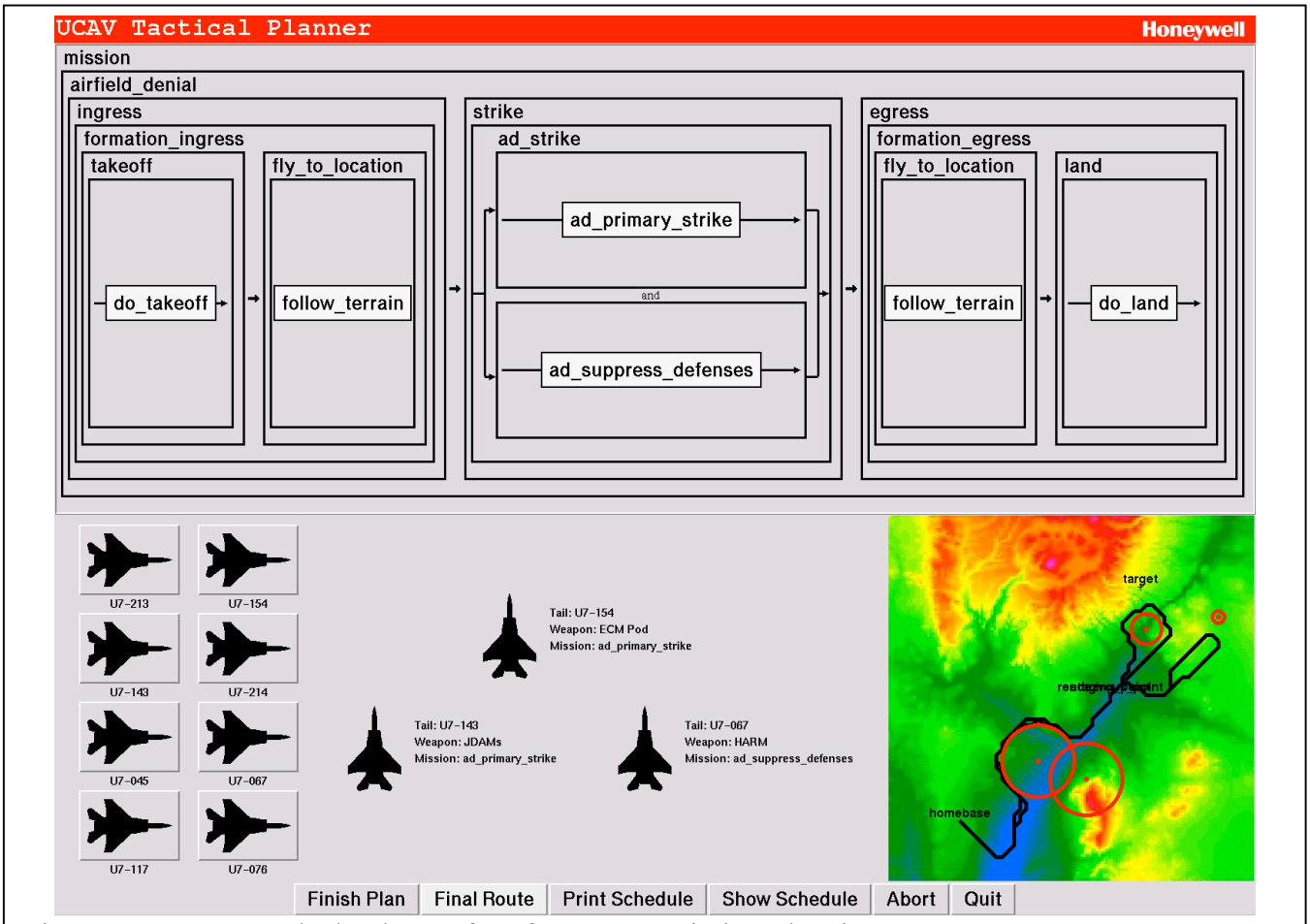


Figure 4. Prototype Playbook Interface for UCAV Mission Planning.

At this point, having been told only that the task for the day is "Airfield Denial," a team of trained pilots would have a very good general picture of the mission they would fly. Similarly, the tasking interface (via the Shared Task Model) knows that a typical airfield denial plan consists of ingress, attack and egress phases and that it may also contain a suppress air defense task before or in parallel with the attack task. But just as a leader instructing a human flight team could not leave the delegation instructions at a simple "Let's do an Airfield Denial mission today," so the operator of the Playbook must provide more information. Here, the user must provide four additional items: a target, a homebase, a staging and a rendezvous point. These items are geographical in nature and users typically find it easier to stipulate them via a map. Hence, by selecting any of them from the menu, direct interaction with the Terrain Map is enabled. Since the Playbook knows what task and parameter a map point is meant to indicate, appropriate semantics are preserved. The specific aircraft to be used may be selected by the user or left to the APC. If the user wishes, available aircraft can be viewed in the Available Resource Space and chosen by clicking and moving them to the Resources in Use Area.

The mission leader working with a team of human pilots could, if time, mission complexity or trust made it desirable, hand the mission planning task off to the team members at this point. The Playbook operator can do this as well, handing the task to the APC via the "Finish Plan" button. The leader might wish, however, to provide more detailed delegation instructions. This can be accomplished by progressively interacting with the Playbook UI to provide deeper layers of task selection, or to impose constraints or stipulations on the resources to be used, waypoints to be flown, etc. For example, after the user chooses "Airfield Denial" the system knows, via the Shared Task Model, that this task must include an Ingress

subtask (as illustrated in Figure 4) and, therefore, the user does not need to tell it this. But, the user can provide detailed instructions about how to perform the Ingress task by simply "drilling down" into it. Choosing "Ingress" produces a "generic" Ingress task template or "play". This need not be a default method of doing "Ingress" but a generic, uninstantiated template—corresponding to what a human expert knows about what constitutes an Ingress task and how it can be performed. A trained pilot knows that Ingress can be done either in formation or in dispersed mode and must involve a "Take Off" subtask followed by one or more "Fly to Location" subtasks. Similarly, the Playbook user can select from available options (e.g., formation vs. dispersed Ingress, altitude constraints, etc.) on context-sensitive, APC-generated menus at each level of decomposition of the task model.

The user can continue to specify and instantiate tasks down to the primitive level where sub-tasks are behaviors the control algorithms (see Figure 3) can be relied upon to execute. Alternatively, at any point after the selection of the top level task and its required parameters, the tasker can hand the partly developed plan over to the APC for completion/review. In extreme cases, a viable "Airfield Denial" plan involving multiple UAVs could be created in our prototype with as few as five selections and more sophisticated planning capabilities could reduce this number further. If the APC is incapable of developing a viable plan within the constraints imposed, (e.g., if the user has stipulated distant targets that exceed aircraft fuel supplies) it will inform the user. The prototype described here gives users limited ability to modify APC-generated plans (only the ability to reject a plan and request a new one with modified constraints), but we realize that full, fielded implementations will require richer "negotiations" about what is and is not feasible, thereby making the planning process more collaborative. We are taking the first steps toward this by enabling the APC to respond with suggested alternatives when the user's request proves infeasible (Goldman, et al., 2005; Miller et al., 2005).

IMPLICATIONS, PRELIMINARY DATA AND FUTURE WORK

The Playbook described above enables a human supervisor to make use of varying LOAs for the task of planning a UAV mission. It accomplishes this flexibility by enabling explicit reference to and manipulation of the levels in a task hierarchy of domain functions. Once referenced, desired actions or constraints can be readily commanded with regards to those functions. Equally important is the presence of an intelligent planning and control environment that understands the human expressions of intent and plans execution within them.

In short, this approach allows "delegation" of the responsibility for planning and commanding a UAV mission in many of the same ways a mission commander might delegate to a team of experienced pilots. Using the task hierarchy, the operator can delegate almost complete autonomy to the automation and receive a complete plan for review. This represents level 8 for the overall task of mission planning in Sheridan's hierarchy (cf. Table 1). It also represents a tradeoff between workload, uncertainty, and competency that is at the right side of the spectrum illustrated in Figure 1. Alternatively, the human can also lay out very detailed instructions, making decisions and stipulating specific actions to complete a mission plan to the primitive level. This represents level 2 or 1 in Table 1, and a tradeoff near the left side of Figure 1.

Most importantly, the human can also interact, on a task by task basis, as desired or needed. If the operator wishes to plan mission ingress in detail, review computer-suggested alternatives for attack, and tell the system that anything it wishes to do for egress is acceptable as long as the UAVs return by 0300 hours—this level of variability is supported. The resulting experienced LOA for the overall mission planning task might be somewhere in the middle of the spectrum—but this is merely a convenient averaging. In fact, the LOA will have varied from task to task at the needs of the operator. Thus, our motivation to decompose a task into its possible components and assign automation

to each of them individually provides not only a theoretical expansion of the LOAs that previous authors have proposed, but also provides the framework for intent communication required to enable complex but flexible, adaptable delegation actions.

Such interactions should result in significant payoffs. Adaptable delegation approaches may mitigate the "ironies" of automation (Bainbridge, 1983) and minimize the risks of traditional automation design as noted at above. User-adaptable delegation approaches may allow the benefits of real-time adaptation while keeping the operator in charge. Since an interaction style allows the user to interact with automation across a range of task abstraction levels, it fulfills the principle of allowing the operator to "finish the design" at the time of use. This should provide users with more control authority, addressing their desire to remain in charge and ameliorating frustration, stress, and lack of user acceptance (Karasek & Theorell, 1990; Miller, 1999; Vicente, 1999). Also, by providing the opportunity for true supervisory control, with all the flexibility of delegation available to human supervisors, we should encourage enhanced human engagement with automation in task performance. This in turn should provide the following benefits:

- Improved situation awareness relative to an interface that automatically provides a solution at a fixed LOA, at least with regards to those information elements that the human is more directly concerned with.
- Improved workload relative to an interface that provides less automation support and requires more human task performance, again with regards to those tasks for which a higher LOA is used.
- Improved overall human + machine performance relative to either human alone or machine alone.

We have begun to demonstrate the feasibility of tasking interfaces in general and Playbook in particular, though much work remains. The early work on adaptive automation (Hancock et al., 1985; Parasuraman et al., 1992; Rouse, 1988) needed validation with empirical studies, which were carried out subsequently (Kaber & Riley, 1999; Parasuraman et al., 1996; for reviews, see Inagaki, 2003; Scerbo, 1996, 2001). Similarly, delegation prototypes need to be evaluated to ascertain both the usability of the specific interface design presented here and the performance benefits of the concept as a whole. We have recently conducted a series of studies investigating some of the claimed benefits using a simplified delegation interface in a human-robot interaction task. Initial results (reported in detail in Parasuraman, Galster, Squire, Furukawa, and Miller, 2005), are promising. Space does not permit detailed reporting here, but these experiments, using a simplified, flexible, delegation interface for a robotic "capture the flag" game, showed that participants chose to make use of flexible LOAs in response to differing environmental conditions. Furthermore, the use of such flexibility was generally associated with the best mission success rates and the shortest mission completion times across varying environmental and "enemy" behavior conditions—though these effects could be nullified when the added workload of controlling large numbers of robots conflicted with the workload of selecting among flexible interaction modes (though even here, performance with the delegation interfaces was as good as the best static or restricted UIs).

The claim that delegation interfaces do not increase workload to unmanageable levels is a critical one that requires validation. Kirlik (1993) reported that workload costs associated with deciding what LOA to employ can overwhelm any payoffs of flexible automation (see also Parasuraman & Riley, 1997). Tasking interfaces, like human-human delegation, require thought about what to delegate and what to retain. Whether human operators (as opposed to designers) can determine the best mix of tasks to delegate in various contexts is perhaps the largest open question in pursuing flexible human supervisory control. For example, Jones and Jasek (1997) report data suggesting that, at least in the domain of satellite communication planning, operators tended to consistently use

the highest LOAs available regardless of context or workload.

We have sought to create a style of interaction with automation that mirrors delegation to an intelligent assistant who knows the task domain. We have adopted a playbook organization using hierarchical task decomposition to facilitate and speed the communication of intent. This is an extraordinarily powerful form of human-human interaction and, as such, has been adopted by military, government, and commercial organizations—not to mention sports teams—throughout history. But just as an inefficient assistant or a poorly trained teammate can require more work than doing the job oneself, so a poorly designed tasking interface may make the task more difficult. Nevertheless, we believe that the delegation concept offers a viable method for flexible, multi-level interaction between humans and automated agents in a manner that enhances system performance while maintaining user workload at a manageable level. We have illustrated one method of constructing such a delegation Playbook for rich, flexible user interactions and are beginning the process of illustrating benefits from it. Future work will continue this process of tuning and evaluating the delegation interaction format.

ACKNOWLEDGEMENTS

The initial formulation of the delegation interface concept and the development of the prototype interface described here was funded by a Honeywell Laboratories Initiatives Grant. The preparation of this paper was funded, in part, by grants from DARPA to R. Parasuraman and C. Miller (subcontract, Cornell University) and by ARL to R. Parasuraman. The authors would like to thank Dan Bugajski, Don Shaner and John Allen for their help in the development and implementation of the ideas presented. Robert Goldman and Michael Pelican were responsible for the specific design and implementation of the example shown in Figure 4. Harry Funk and Dave Musliner participated in the initial development of these concepts. We also thank Scott Galster and Peter Squire for helpful discussions on the use of Playbook interfaces.

REFERENCES

Amalberti, R. (1999). Automation in aviation: A human factors perspective. In D. Garland, J. Wise & V. Hopkin (Eds.), *Handbook of aviation human factors* (pp. 173-192). Mahwah, NJ: Erlbaum.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*, 775-779.

Banks, S. & Lizza, C. (1991). Pilots' Associate; A cooperative knowledge-based system application, *IEEE Expert, vol. 6(3),* 18-29.

Billings, C. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Erlbaum.

Billings, C. E. and Woods, D. D. 1994, Concerns about adaptive automation in aviation systems, in M. Mouloua and R. Parasuraman (eds), *Human performance in automated systems: Current research and trends*. Hillsdale, NJ: Lawrence Erlbaum. 24–29.

Bisantz, A., Cohen, S., Gravelle, M., & Wilson, K. (1996). "To cook or not to cook: A case study of decision aiding in quick-service restaurant environments," Georgia Institute of Technology, College of Computing, Atlanta, GA, Report No. GIT-CS-96/03.

Bright, J. (1955). Does automation raise skill requirements? *Harvard Business Review, 36*, 85-98.

Card, S., Moran, T., & Newell, A. (1983) *The psychology of human computer interaction,* Mawah, NJ; Lawrence Erlbaum Associates.

Casey, S. (1993). *Set phasers on stun*, Santa Barbara, CA: Aegean Publishing Company.

Colucci, F. (1995, March/April). Rotorcraft Pilots' Associate update: The Army's largest science and technology program, *Vertiflite*, 16-20.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. (2000). Measuring team knowledge. *Human Factors, 42*, 151-173.

Degani, A. (2004). *Taming Hal: Designing interfaces beyond 2001*. New York: Palgrave MacMillan.

Diaper, D. (1989). *Task analysis for human-computer interaction*, Chichester, UK: Ellis Horwood.

Dornheim, M. (1999). Apache tests power of new cockpit tool. *Aviation Week and Space Technology*, October 18, pp. 46-49.

Duley, J. & Parasuraman, R. (1999) Adaptive information management in future air traffic control, In M. Scerbo & M. Mouloua, (Eds.), *Automation technology and human performance: Current research and trends,* (pp.86-90). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. & Kiris, E. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*. 381-394.

Erol, K., Hendler, J., & Nau, D. (1994). UMCP: A sound and complete procedure for hierarchical task network planning," in *AI Planning Systems: Proceedings of the 2nd International Conference,* K. Hammond, Ed., Los Altos, CA, pp. 249-254.

Geddes, N. (1989). Understanding human operators' intentions

in complex systems. (Doctoral Dissertation, Georgia Institute of Technology, 1989).

Gentner, D. & Stevens, A. (1986). *Mental Models*. Hillsdale, NJ; Lawrence Erlbaum Associates.

Goldman, R., Haigh, K., Musliner, D., & Pelican, M. (2000). MACBeth; A multi-agent, constraint-based planner. *Notes of the AAAI Workshop on Constraints and AI Planning*, Austin, TX, pp. 1-7.

Goldman, R., Miller, C., Wu, P., Funk, H. & Meisner, J. (2005). Optimizing to Satisfice: Using Optimization to Guide Users. In *Proceedings of the American Helicopter Society's International Specialists Meeting on Unmanned Aerial Vehicles*. January 18-20; Chandler, AZ.

Guerlain, S. (1995). Using the critiquing approach to cope with brittle expert systems. In *Procedings of the Human Factors and Ergonomics Society, 39th Annual Meeting,* Santa Monica, CA, pp. 233-237.

Hancock, P., Chignell, M., & Lowenthal, A. (1985). An adaptive human-machine system. *Proceedings of the IEEE Conference on Systems, Man and Cybernetics, 15*, pp. 627-629.

Hentenryck, P. (1989) *Constraint satisfaction in logic programming.* Cambridge, MA: MIT Press.

Hilburn, B., Jorna, P., Byrne, E., & Parasuraman, R. (1997) The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. Koonce (Eds.), *Human-automation interaction* (pp.84-91). Mahwah, NJ: Erlbaum.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. In E. Hollnagel (Ed.), *Handbook of cognitive task design*. Mahwah, N.J.: Lawrence Erlbaum Publishers.

Jaffar, J. & Michaylov, S. (1987). Methodology and implementation of a CLP system. In *Proceedings of the 4th International Conference on Logic Programming*. Cambridge, MA: MIT Press.

Jones, P. & Jasek, C. (1997). Intelligent Support for Activity Management (ISAM): An architecture to support distributed supervisory control. *IEEE Transactions on Systems, Man & Cybernetics—Pt. A: Systems and Humans*, *27*, 274-288.

Kaber, D., Omal, E. & Endsley, M. (1999). Level of automation effects on telerobot performance and human operator situation awareness and subjective workload. In M. Mouloua & R. Parasuraman, (Eds.), *Automation technology and human performance: Current research and trends*, (pp. 165-170). Mahwah, NJ: Erlbaum.

Kaber, D. B., & Endsley, M. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theoretical Issues in Ergonomics Science, 5, 113-153.

Kaber, D., & Riley, J. (1999) Adaptive automation of a dynamic control task based on workload assessment through a secondary monitoring task. In M. Scerbo & M. Mouloua, (Eds.), *Automation technology and human performance: Current research and trends,* (pp.129-133). Mahwah, NJ: Lawrence Erlbaum.

Karasek, R. & Theorell, T. (1990). *Healthy work: Stress, productivity, and the reconstruction of working life*. New York: Basic Books.

Keller, K & Stanley, K. (1992). Domain specific software design for decision aiding, *Proc. AAAI Workshop on Automating Software Design,* NASA Ames Research Center, pp. 86-92.

Kirlik, A. (1993). Modeling strategic behavior in human utomation interaction: Why an 'aid' can (and should) go unused. *Human Factors, 35*, 221-242.

Kirwan, B. & Ainsworth, L. (1999). *A Guide to Task Analysis*. London: Taylor & Francis.

Klein, G. (1998). *Sources of power; How people make decisions.* Cambridge, MA:MIT Press.

Layton, C., Smith, P., & McCoy, E. (1994). Design of a cooperative problem solving system for enroute flight planning: An empirical evaluation. *Human Factors, 36(1),* 94-119.

Lee, J., & Moray, N. (1992). Trust, control strategies, and allocation of function in human-machine systems. *Ergonomics*, *35*. 1243-1270.

Lee, J., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40,* 153-184.

Lee, J. D. & See, K. A. (2004). Trust in computer technology: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80.

Lewis, M. (1998). Designing for human-agent interaction. *Artificial Intelligence Magazine,* (Summer), 67-78.

Lorenz, B., Nocera, F., Rottger, S., & Parasuraman, R. (2002). Automated fault management in a simulated space flight micro-world. *Aviation, Space and Environmental Medicine, 73*. 886-897.

Martinich, J. (1996). *Production and operations management: An applied modern approach,* New York; John Wiley.

Maybury, M. & Wahlster, W. (1998). *Readings in Intelligent User Interfaces*. San Francisco, CA; Morgan Kaufman.

Milewski, A. & Lewis, S. (1999) When People Delegate. AT&T Technical Memorandum.

Miller, C. (1999). Bridging the Information Transfer Gap: Measuring Goodness of Information Fit. *Journal of Visual Language and Computation, 10,* 523-558.

Miller, C., Goldman, R., Funk, H., Wu, P. & Pate, B. (2004). A Playbook Approach to Variable Autonomy Control: Application for Control of Multiple, Heterogeneous Unmanned Air Vehicles. In Pro*ceedings of FORUM 60, the Annual Meeting of the American Helicopter Society*. Baltimore, MD; June 7-10.

Miller, C. & Hannen, M. (1999). The Rotorcraft Pilot's Associate: Design and evaluation of an intelligent user interface for cockpit information management. *Knowledge Based Systems*, *12,* 443-456.

Mitchell, C. (1987). GT-MSOCC: A domain for research on HCI and decision aiding in supervisory control systems. *IEEE Trans. on Systems, Man and Cybernetics, 17(4),* 553-572.

Moray, N. (1979). *Mental workload.* New York: Plenum.

Moray, N., Inagaki, T., & Itoh, M. (2000). Situation adaptive automation, trust and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied, 6*(1), 44-58.

Murata, T. (1989). Petri Nets: Properties, analysis and applications, *Proceedings of the IEEE, 77(4)*, 541-580.

Musliner, D., Durfee, E., and Shin, K. (1993). CIRCA: A Cooperative Intelligent Real-Time Control Architecture. *IEEE Transactions on Systems, Man, and Cybernetics, 23*(6). 1561-1574.

Nikolic, M.I, & Sarter, N.B. (2000). Peripheral visual feedback: A powerful means of supporting attention allocation and human-automation coordination in highly dynamic data-rich environments. *Human Factors, 43(1)*, 30-38.

Opperman, R., (1994). *Adaptive user support.* Hillsdale, NJ; Erlbaum.

Pande, A. & Hayes, C. (2002). Plan-Edit: An interactive critic for user guided generation of high quality manufacturing

setup sequences. *Proceedings of ASME International Congress and Exposition*, Montreal, Canada.

Parasuraman, R. (1993). Effects of adaptive function allocation on human performance. In D. J. Garland & J. A. Wise (Eds.), *Human factors and advanced aviation technologies* (pp. 147-157). Daytona Beach: Embry-Riddle Aeronautical University Press.

Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics, 43*, 931-951.

Parasuraman, R., Bahri, T. Deaton, J., Morrison, J. & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems.* (Progress Report No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center.

Parasuraman, R., Galster, S., Squire, P., Furukawa, H. & Miller, C. (2005). A Flexible Delegation-Type Interface Enhances System Performance in Human Supervision of Multiple Robots: Empirical Studies with RoboFlag. *IEEE Systems, Man and Cybernetics—Part A: Systems and Humans, 35*(4), 481-493.

Parasuraman, R., & Hancock, P. (2001). Adaptive control of workload. In P. A. Hancock, & P. E. Desmond (Eds.), *Stress, workload, and fatigue* (pp.305-320). Mahwah, NJ: Erlbaum.

Parasuraman, R. & Miller, C. A. (in press). Delegation interfaces for human supervision of multiple unmanned vehicles: Theory, experiments, and practical applications. In N. Cooke, N. and H. Pedersen (Eds.) *Human Factors of Remotely Piloted Vehicles*.

Parasuraman, R. & Mouloua, M., Eds. (1996). *Automation and Human Performance: Theory and Application* (pp. 1-16). Mahwah, NJ: Erlbaum.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. W. Scerbo & M. Mouloua (Eds.) *Automation technology and human performance: Current research and trends.* (pp. 119-123). Mahwah, NJ: Erlbaum.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, *38,* 665-679.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

Parasuraman, R., Sheridan, T. & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, *30*, 286-297.

Prinzel, L., Freeman, F., Scerbo, M., Mikulka, P. & Pope, A. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload and event-related potential P300 component. *Human Factors, 45(4*), 601-614.

Perrow, C. (1984). *Normal accidents: Living with high-risk technologies.* USA: Basic Books.

Rasmussen, J. (1986). *Information processing and human-machine interaction.* Amsterdam: North-Holland.

Rasmussen, J., Pejtersen, A. & Goodstein, L. (1994). *Cognitive Systems Engineering*. New Yorkl; John Wiley & Sons, Inc.

Reason, J. (1997). *Managing the risks of organizational accidents.* Aldershot, UK: Ashgate Publishing Ltd.

Rouse, W. (1976). Adaptive allocation of decision making responsibility between supervisor and computer. In T.B. Sheridan & G. Johannsen, (Eds.), *Monitoring behavior and supervisory control* (pp.221-230). New York: Plenum Press.

Rouse, W. (1988). Adaptive automation for human/computer control. *Human Factors*, *30*, 431-488.

Sarter, N. & Woods, D (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *Int. Jour. of Av. Psych., 4*, 1-28.

Sarter, N. and Woods, D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, *37*(1), 5-19.

Sarter, N., Woods, D. & Billings, C. (1997). Automation surprises. In G. Salvendy, (Ed.) *Handbook of human factors and ergonomics*, *2nd ed*. (pp.1926-1943). New York: Wiley.

Satchell, P. (1998). *Innovation and automation*. Aldershot, UK: Ashgate.

Scallen, S., Hancock, P., & Duley, J. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergo*nomics, *26*. 397-403.

Scerbo, M. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman, & M. Mouloua (Eds.), *Automation & human performance: Theory & applications* (pp.37-63). Mahwah, NJ: Erlbaum.

Scerbo, M. (2001). Adaptive automation. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 1077-1079). London: Taylor & Francis.

Shepherd, A. (1989). Analysis and training in information technology tasks. In D. Diaper (Ed.) *Task Analysis for Human-Computer Interaction*. Chichester: Ellis Horwood, Ltd. 15-55.

Shattuck, L. (1995). *Communication of intent in distributed supervisory control systems*. Unpublished doctoral dissertation, The Ohio State University, Columbus, OH.

Sheridan, T. (1987). Supervisory control. In G. Salvendy, (Ed*.), Handbook of human factors* (pp.1244-1268). New York: John Wiley & Sons.

Sheridan, T. (1992). *Telerobotics, automation, and supervisory control*. Cambridge, MA: MIT Press.

Sheridan, T., (2002). *Humans and automation*. Santa Monica and NY: Human Factors and Ergonomics Society & Wiley.

Sheridan, T., & Verplank, W. (1978). *Human and Computer control of undersea teleoperators*. MIT Man-Machine Systems Laboratory, Cambridge, MA. Technical Report.

Stout, R.,J., Cannon-Bowers, J.A., Salas, E., & Milanovich, D.M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. Human Factors, 41, 61-71.

Suchman, L. (1987). *Plans and situated actions; The problem of human machine communication*. Cambridge; Cambridge University Press.

Vicente, K. (1999). *Cognitive work analysis: Towards safe, productive, and healthy computer-based work*. Mahwah, NJ; Erlbaum.

Vicente, K. (2003). *The Human Factor*. New York: Alfred Knopf

Wickens, C. (1994). Designing for situation awareness and trust in automation. In *Proceedings of the IFAC Conference.* Baden-Baden, Germany, pp. 174-179.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance.* New York: Longman.

Wickens, C., Mavor, A., Parasuraman, R. & McGee, J. (1998). *The future of air traffic control: Human operators and automation.* Washington DC: National Academy Press.

Wiener, E. (1988). Cockpit automation. In E. Wiener & D.

Nagel, (Eds.), *Human factors in aviation* (pp.433-461). San Diego: Academic.
Wiener, E., & Curry, R. (1980). Flight-deck automation: Promises and problems. *Ergonomics, 23*, 995-1011.
Woods, D. (1996). Decomposing automation: Apparent simplicity, real complexity. In
Parasuraman, R., Mouloua, M., & Molloy, R. (Eds.), *Automaton and human performance: Theory and applications* (pp.3-17). Mahwah, NJ: Earlbaum.

Christopher A. Miller received his Ph.D. from the Committee on Cognition and Communication in the Psychology Department of the University of Chicago in 1991. He is now the Chief Scientist at Smart Information Flow Technologies, LLC. based in Minneapolis, MN, and can be reached at cmiller@sift.info.

Raja Parasuraman received his Ph.D. in Psychology from Aston University, U.K. in 1976. He is Professor of Psychology and a member of the Arch Lab at George Mason University, Fairfax, VA. He can be reached at rparasur@gmu.edu.

***Human Factors*, in press**

Effects of Imperfect Automation on Decision-Making in a Simulated Command and Control Task

ERICKA ROVIRA[1], KATHLEEN MCGARRY[1], & RAJA PARASURAMAN[2]

The Catholic University of America, Washington, DC, USA[1]

George Mason University, Fairfax, VA, USA[2]

Objective. Effects of four types of automation support and two levels of automation reliability were examined. The objective was to examine the differential impact of information and decision automation and to investigate the costs of automation unreliability. *Background.* Research has shown that imperfect automation can lead to differential effects of stages and levels of automation on human performance. *Method.* Eighteen participants performed a "Sensor to Shooter" targeting simulation of command and control. Dependent variables included accuracy and response time of target engagement decisions, secondary task performance, and subjective ratings of mental workload, trust, and self-confidence. *Results.* Compared to manual performance, reliable automation significantly reduced decision times. Unreliable automation led to greater cost in decision-making accuracy under the higher automation reliability condition for three different forms of decision automation relative to information automation. At low automation reliability, however, there was a cost in performance for both information and decision automation. *Conclusion.* The results are consistent with a model of human-automation interaction that requires evaluation of the different stages of information processing to which automation support can be applied. *Application.* If fully reliable decision automation cannot be guaranteed, designers should provide users with information automation support or other tools that allow for inspection and analysis of raw data.

Reprints: Ericka Rovira
United States Military Academy
Department of Behavioral Sciences and Leadership
Building 601, Cullum Road
West Point, NY 10096
rovira.ericka@gmail.com

Running head: Automation in Command and Control

Key Words: automation, command and control, complacency, decision-making, decision support

INTRODUCTION

Military command and control (C^2) typically requires strategic and tactical decisions to be made in a timely manner based on multiple sources of information. A major goal for C^2 systems is to shorten the targeting cycle or tighten the *sensor to shooter* loop (i.e. execute the process more quickly than the enemy) (Adams, 2001). The sensor to shooter cycle includes processes from the time a sensor spots a target until a shooter locks on it. Shortening the cycle by necessity requires that the human operators of C^2 systems be supported in some manner. Automation is one major form of support.

An important design issue for automated support tools is how much authority to assign to the automation. Factors influencing this decision include how well the automation supports the operator's situation awareness as well as the type and reliability of the automation (Barnes, 2003). Unfortunately, automated aids have not always enhanced system performance, primarily due to problems in their use by human operators or to unanticipated interactions with other sub-systems. Problems in human-automation interaction have included unbalanced workload, reduced system awareness, decision biases, mistrust, over-reliance, and complacency (Parasuraman & Riley, 1997). There is thus a need for designing automation that supports military operators in command and control in ways that avoid such negative influences.

Parasuraman, Sheridan, and Wickens (2000) proposed a taxonomy that can guide automation design. They identified four stages of human information processing which may be supported by automation: (stage 1) information acquisition; (stage 2) information analysis; (stage 3) decision and action selection; and (stage 4) action implementation. Each of these stages may be supported by automation to varying degrees, between the extremes of manual performance and full automation (Sheridan & Verplank, 1978). Because they deal with distinct aspects of information processing, the first two (information acquisition and analysis) and the last two stages (decision selection and action implementation) are sometimes grouped together and referred to as *information* and *decision* automation, respectively (see also Billings, 1997; Lee & Sanquist, 1996). Information automation can assist the user in inference and recommend possible courses of action (Jones et al., 2000). Decision automation involves selecting from various decision options and is generally based on an inference of the state of the world (information automation); however different outcomes may be given certain values, hence making automation of this stage distinct from information automation (Wickens, 2000).

The sensor to shooter loop can be directly mapped to the Parasuraman et al. (2000) model: target information is first acquired from sensors, subsequently analyzed further before a course of action is decided on, and finally acted upon. One example of a fielded C^2 system in which automation is applied to different stages and at different levels is the Theater High Altitude Area Defense (THAAD) system. THAAD, a system used to intercept ballistic missiles (Department of the Army,

2003), has relatively high levels of information acquisition, information analysis, and decision selection; however action implementation automation is low, giving the human control over the execution of a specific action.

Several studies have examined the differential effects of stages and levels of automation on human performance (Crocoll & Coury, 1990; Endsley & Kaber, 1999; Galster, Bolia, & Parasuraman, 2002; Lorenz, Di Nocera, Röttger, & Parasuraman, 2002; Sarter & Schroeder, 2001; Wickens & Xu, 2002). Crocoll and Coury (1990) examined three versions of an automated decision aid in an air defense targeting (identification and engagement) task. The automation provided (1) status information about a target (information automation), (2) recommendation concerning its identification (decision automation), or (3) both (information + decision automation). Crocoll and Coury found that with an imperfect aid, participants were better able to recover when provided with only status information. Sarter and Schroeder (2001) conducted a similar study involving automated decision aiding in a flight simulation environment. The decision aid provided status aiding (information automation) or command aiding (decision automation) regarding in-flight icing. The results confirmed the findings of Crocoll and Coury (1990) that imperfect automation can lead to worse performance when it provides decision support than when it gives only information support.

The present study compared information and decision automation, but differed in the following. First, we investigated automated aiding in a command and control environment to examine the replicability and generality of the previous results to a different simulated task domain. Second, the level of automation unreliability was fixed in the previous studies. We varied automation unreliability over two levels, allowing an assessment of the range of automation unreliability effects on decision-making, as well as a comparison to previous work on automation complacency in which automation reliability was varied and operator reliance on automated aiding was evaluated (May, Molloy, & Parasuraman, 1993; Parasuraman, Molloy, & Singh, 1993). Our hypothesis for examining this factor was guided by one of the paradoxes of automation — that when automation is imperfect, *greater* imperfection may lead to a lowering of the cost of unreliability. We therefore predicted that the differential cost of unreliable decision automation would be greater for high than for low automation reliability. Third, we developed three different kinds of decision automation to examine the generality of the phenomenon of the unreliability of decision automation. Lastly, in the present study we examined the *first automation failure effect*, which refers to the influence on user performance of the initial failure of previously perfect automation. The assumption is that while operators may be complacent as a result of exposure to apparently perfect automation and not respond appropriately to the first occurrence of a failure, further experience with imperfect automation may lead to appropriate calibration of trust and subsequently improved performance. In an examination of this effect, Merlo, Wickens,

and Yeh (2000) measured target detection times in a cued search task and introduced several cueing errors in the final experimental block. Participants improved their detection of the uncued high priority target after experiencing the first cueing error (accuracy after was 91% versus 50% before), suggesting that participants used a different strategy after some exposure to imperfect automation, thus improving their performance on later trials. Consequently, one additional objective of our study was to examine first automation failure effects with information and decision automation. We hypothesized that the first time automation failed operators would perform poorly, but that after some experience they would be able to calibrate their trust accordingly, thus reducing complacency and improving performance.

METHODS

Participants

Eighteen undergraduate students (nine males and nine females) aged 18 to 22 years (M = 19.83, SD = 1.25) from The Catholic University of America volunteered and were paid $15 per hour for their participation.

Apparatus: The Sensor to Shooter Task Environment

A PC-based low-fidelity software simulation of a sensor to shooter targeting system (STS) was developed. The program was written in Java code for a Pentium-II processor with a 17-inch monitor. A mouse was used as the input device. The STS simulation consisted of three components shown in separate windows: a terrain view, a task window, and a communications module. The right portion of the screen was dedicated to a two-dimensional terrain view of a simulated battlefield. The window showed three red enemy units (labeled E1, E2, and E3), three yellow friendly battalion units (B1, B2, and B3), and six green friendly artillery units (A1, A2, A3, A4, A5, and A6), as well as one blue friendly headquarter unit (HQ). A second window, the task window, was where the user made enemy-friendly engagement selections. The participants were required to identify the most dangerous enemy target and to select a corresponding friendly unit to engage in combat with the target. Figure 1 shows a static view of the STS simulation.

Insert Figure 1 about here

The bottom left portion of the task window was allocated to the automation support. (1) *Information automation* provided the participant with a complete list of all possible engagement combinations including the distances between enemy targets, friendly units, and headquarters. Because no explicit pointer to decision selection was provided, this corresponds to information (stage 2) automation in the Parasuraman et al. (2000) taxonomy. The remaining types represented different varieties of decision support: (2) *Low decision automation* gave a complete list of all possible engagement combinations including the distances between enemy targets, friendly units, and headquarters; however, in this instance the listings were

prioritized with the best selection first and the worst choice last, thereby making this a form of decision automation (stage 3). (3) In the *medium decision automation* condition, the participant was provided the top three options for engagement, including the distances between enemy targets, friendly units, and headquarters. (4) The *high decision automation* condition recommended the best enemy-friendly engagement. Although distance information between enemy targets, friendly units, and headquarters was computed by the automation algorithm, it was not readily accessible to the operator. Thus the highest form of decision automation removed some of the "raw" data.

The communications task window, a secondary task, was located in the top left portion of the screen. A call sign appeared every five seconds and remained displayed until the next call sign. A call for communications was randomly distributed once over a 50 second time period.

Task Procedures and Design

An enemy-friendly engagement selection was required within 10 s. Additionally, participants were required to click on the call answer button every time their call sign appeared in the communications window.

A 4 (type of automation) x 2 (overall automation reliability) x 2 (trial reliability) within-subjects design was used. The four automation support conditions included information automation, and three different forms of decision automation: low, medium, and high. Overall automation reliability was varied across two values (60% and 80%) by manipulating the proportion of trials in which a correct assessment or decision was provided. In the 60% overall automation reliability condition, 24 trials were reliable (correct automated assessment) and 16 trials were unreliable (incorrect automated assessment). Similarly, in the 80% overall automation reliability condition, 32 trials were reliable and eight trials were unreliable. Reliability was manipulated for each of the four automation support conditions. Trial reliability referred to a correct automated assessment (reliable) versus an incorrect automated assessment (unreliable). Participants were informed that while the automation was highly reliable, it was not 100% reliable all of the time. However, no further information on the actual reliability was given.

Each participant first performed the task manually (20 trials) with no automation support. Each participant experienced 40 trials of each of the 8 conditions: information automation at (1) 60% and (2) 80% reliability; low decision automation at (3) 60% and (4) 80% reliability; medium decision automation, at (5) 60% and (6) 80% reliability; and high decision automation at (7) 60% and (8) 80% reliability. The order of the four automation support tools was counterbalanced between participants. Participants completed both overall automation reliability conditions (60% and 80%) for each automation support tool before a new automation support tool was introduced. Overall automation reliability was

counterbalanced within each type of automation. Prior to the start of each automation condition, participants completed 20 trials of 100% reliability automation; these trials were given to allow operators to recover trust in the automation, but were not part of the data analysis. In all, each participant completed 500 trials (20 manual trials, 320 experimental automation trials, and 160 "recovery" (100% reliability) trials.

Dependent variables included the accuracy and speed of enemy-friendly engagement selections. Accuracy was calculated by the percentage of trials in which the participant correctly selected the most dangerous enemy target and a corresponding friendly unit to engage in combat. Secondary task measures of performance included accuracy and response time to the communications (call sign) task. To obtain subjective measures of mental workload, participants were asked to fill out the NASA-TLX (Hart & Staveland, 1988). Subjective ratings of trust in automation and operator self-confidence were obtained on a Likert-rating scale, ranging from 0-100, built after scales used by Lee and Moray (1994). All subjective measures were administered after each automation condition.

RESULTS

Manual Control vs. Automation Support

Decision-Making Accuracy. The mean rate of correct enemy-friendly engagement selections (decision-making accuracy) was computed for each participant under both manual control and with automation support. Under imperfect automation (60% and 80% reliability), engagement selection rates were computed separately for reliable and unreliable trials within each condition. A paired samples *t*-test of engagement selection rates showed that there was no difference in decision accuracy between manual (M = 89.4%) and reliable automation (M = 88.4%), $t(17) = .62$, $p = .541$. However, there was a significant difference in decision accuracy between the manual and the unreliable automation support conditions, $t(17) = 6.9$, $p < .001$, with accuracy declining to 70% under unreliable automation. In general, there was no difference in accuracy performance between manual and reliable automation, but accuracy declined under unreliable automation.

Response Times. Enemy-friendly engagement response times (RTs) were reduced significantly from the manual condition when participants were given reliable automation support, $t(17) = 5.3$, $p < .001$. Under the highest form of decision automation, participants only had to click “Ok” if they agreed with the automation, and this may have lead to a difference in RTs due simply to the design of the interface and the requirement to make a single motor response, rather than to the functionality of the automation. Therefore, a paired samples *t*-test of engagement selections between manual and reliable automation excluding the trials with the highest form of decision automation was conducted. This revealed that a reduction

in RTs still occurred due to reliable automation, $t(17) = 2.37$, $p = .030$. Although RTs increased under unreliable automation, this increase was not significantly different from the manual condition, $t(17) = -.49$, $p = .633$.

Effects of Automation Support and Reliability on Decision-Making Accuracy

The mean rate of correct enemy-friendly engagement selections (decision accuracy) was computed for each participant. A 4 (type of automation: information, low, medium, or high decision automation) x 2 (overall automation reliability: 60% or 80%) x 2 (trial reliability: reliable or unreliable) repeated measures ANOVA was used to examine decision accuracy. The main effects of type of automation, $F(3, 51) = 2.13$, $p = .108$, and of overall automation reliability were not significant, $F(1, 17) = 1.32$, $p = .267$, but that of trial reliability was, $F(1,17) = 34.54$, $p < .001$. Mean accuracy rates for reliable and unreliable trials were 88.5% and 70.0%, respectively. The interaction between type of automation and overall automation reliability was significant, $F(3,51) = 11.18$, $p < .001$. Neither the interaction between automation type and trial reliability, $F(3,51) = 1.34$, $p = .273$, nor that between overall automation reliability and trial reliability were significant, $F(1,17) = .46$, $p = .508$. However, there was a significant three-way interaction between type of automation, overall automation reliability, and trial reliability, $F(3,51) = 13.962$, $p < .001$. To examine these effects more closely, individual comparisons between levels of trial reliability (reliable vs. unreliable) across all types of automation support (information automation, low, medium, and high decision automation) and levels of overall automation reliability (60% and 80%) were conducted.

For the 80% overall automation reliability conditions, the effect of trial reliability was not significant for information automation, $t(17) = .141$, $p = .89$, but was for each of the three decision automation conditions. Accuracy was higher on reliable trials than on unreliable trials for low ($M = 88.62\%$ vs. $M = 66.4\%$), $t(17) = -2.78$, $p = .013$, medium ($M = 86.14\%$ vs. $M = 59.17\%$), $t(17) = -3.47$, $p = .003$, and high decision automation ($M = 95.72\%$ vs. 64.67%), $t(17) = -4.01$, $p = .001$. As Figure 2A shows, trial unreliability did not degrade performance in the information automation condition but did for all three decision automation conditions. As shown in Figure 2B however, for the 60% overall automation reliability condition, there was a cost of trial unreliability for information automation ($M = 83.51\%$ reliable, 53.06% unreliable), $t(17) = -7.16$, $p < .01$, and two of the three decision automation conditions: medium decision automation ($M = 91.44\%$ reliable, 72.97% unreliable), $t(17) = -3.15$, $p = .006$, and high decision automation ($M = 89.64\%$ reliable, 73.39% unreliable), $t(17) = -2.62$, $p = .018$. The effect of trial reliability was not significant for low decision automation, $t(17) = -.97$, $p = .347$.

Insert Figure 2A and 2B about here

Effects of Automation Support and Reliability on Response Time

Enemy-friendly engagement response times were examined using a 4 (type of automation: information, low, medium, or high decision automation) x 2 (overall level of automation reliability: 60% or 80%) x 2 (type of trial reliability: reliable or unreliable) repeated measures ANOVA. The main effect of type of automation was significant, $F(3, 51) = 8.61$, $p < .001$. The mean enemy-friendly engagement response times for information automation, low, medium, and high decision automation were 6037 ms, 5795 ms, 6172 ms, and 5210 ms respectively. The main effect of overall automation reliability was marginally significant, $F(1, 17) = 3.33$, $p = .086$, the mean enemy-friendly engagement response times for 60% and 80% reliability being 5904 ms and 5703 ms respectively. The main effect of trial reliability was significant, $F(1,17) = 43.65$, $p < .001$, with mean enemy-friendly engagement response times for reliable and unreliable trials being 5382 ms and 6225 ms, respectively. The interaction between automation type and overall automation reliability was significant, $F(3,51) = 9.38$, $p < .001$. The interaction between automation type and trial reliability, $F(3,51) = 17.52$, $p < .001$, and the three-way interaction between automation type, overall automation reliability, and trial reliability, $F(3,51) = 2.99$, $p = .04$, were significant.

For the 80% overall automation conditions, the effect of trial reliability was not significant for the information automation condition, $t(17) = -1.52$, $p = .147$, but was significant for two of the three decision automation conditions. The effect of trial reliability was significant for low decision automation, $t(17) = 3.18$, $p = .005$, with shorter response times on reliable trials ($M = 5881$ ms) than on unreliable trials ($M = 6719$ ms). The effect of trial reliability was also significant for high decision automation ($M = 3689$ ms reliable, $M = 6059$ ms unreliable), $t(17) = 6.13$, $p < .001$, but only marginally significant for medium decision automation ($M = 6045$ ms reliable, $M = 6691$ ms unreliable), $t(17) = 1.85$, $p = .081$. For the 60% overall automation reliability condition, however, there was a cost of trial unreliability for information automation ($M = 5794$ ms reliable, $M = 6203$ ms unreliable), $t(17) = 3.11$, $p = .006$; and two of the three decision automation conditions: medium decision automation ($M = 5515$ ms reliable, $M = 6437$ ms unreliable), $t(17) = 5.03$, $p < .001$, and high decision automation ($M = 4450$ ms reliable, $M = 6540$ ms unreliable), $t(17) = 6.34$, $p < .001$. The effect of trial reliability was not significant for low decision automation, $t(17) = .54$, $p = .596$.

First Failure Analysis

The mean decision-making accuracy across participants was first calculated for each automation failure in the medium and high decision automation conditions. We then plotted the data to assess each participant's engagement decision following the first failure in each condition. We found that only two participants made correct decision engagement responses to all first failures in each automated system. Further, we found that regardless of whether participants made

correct or incorrect decision responses to the initial first failure in the experiment, they still made incorrect decision responses to first failures in subsequent automated conditions. Using the binomial distribution, we determined estimates of the mean, π, and standard deviation, $\sigma = \sqrt{[\pi(1-\pi/N)]}$ of the probability of making a correct decision following the first automation failure (Ott, 1993), where π = number of participants making a correct decision/total number of participants, *N*, on that automation failure trial. Next, we used π and σ to compute the 90% confidence intervals for correct response to the first failure and then examined which of the correct response rates to subsequent failures lay outside the interval. If the correct response rate for a particular failure number was above the upper interval value, one can conclude that decision accuracy improved compared to the first failure. For this analysis, we collapsed across automation reliability conditions (60% and 80%). Failure numbers 4, 5, and 8 showed significant improvement over the first failure correct response rate with high decision automation support, whereas only failure number 7 did with medium decision automation support. There was no consistent first failure effect with either low decision automation or information automation support. There appeared to be first failure effects for 2 out of our 3 decision automation support tools. Figure 3 shows clear trends for the first failure effect for both high decision and medium decision automation.

Insert Figure 3 about here

Secondary Task Performance (Communications)

Accuracy. A 4 (type of automation: information, low, medium, or high decision automation) x 2 (overall automation reliability: 60% or 80%) repeated measures ANOVA was used to examine communication task accuracy rates. The only significant effect was the interaction between automation type and overall automation reliability, $F(3,51) = 6.07$, $p < .001$. The effect of overall automation reliability for information automation, $t(17) = .14$, $p = .888$, and low decision automation was not significant, $t(17) = .68$, $p = .506$, but was for medium decision automation, $t(17) = -4.99$, $p = .001$. Mean accuracy rates indicated better performance under low (M = 89.7%) than under high (M = 83.6%) overall automation reliability. The effect of overall automation reliability in the high decision condition was also significant, $t(17) = 2.60$, $p = .019$, and showed accuracy to be better under high (M = 89.1%) than under low (M = 83.9%) overall automation reliability.

Response Times. For RTs in the communications task, a 4 (type of automation: information, low, medium, or high decision automation) x 2 (overall automation reliability: 60% or 80%) repeated measures ANOVA was used. The main effect of type of automation, $F(3, 51) = 7.87$, $p < .001$), was significant with high decision automation leading to lower RTs than the other three conditions. No other effects were significant.

Subjective Ratings of Mental Workload

Using a 4 (type of automation: information, low, medium, or high decision automation) x 2 (overall automation reliability: 60% or 80%) repeated measures ANOVA there was a significant effect of type of automation, $F(3,51) = 4.66$, $p = .006$. Subjective ratings of workload were highest for information automation ($M = 38$), followed by medium decision ($M = 36$), high decision ($M = 34$), and low decision automation ($M = 29$). The effect of overall automation reliability on subjective workload ratings was also significant, $F(1,17) = 7.89$, $p < .012$, with higher ratings for the 80% overall automation reliability condition ($M = 35$) than the 60% overall automation reliability condition ($M = 33$).

Subjective Ratings of Trust and Self-Confidence

A 4 (type of automation: information, low, medium, or high decision automation) x 2 (overall automation reliability: 60% or 80%) repeated measures ANOVA was used to examine subjective ratings of trust. There was a significant effect of type of automation on ratings of trust, $F(3,51) = 3.39$, $p = .025$. Subjective ratings of trust were highest for low decision ($M = 53$), followed by high decision ($M = 48$), medium decision ($M = 41$), and information automation ($M = 34$). Unexpectedly, overall automation reliability had no significant effect on trust, $F(1,17) = .14$, $p = .710$. The interaction between type of automation and overall automation reliability was also not significant, $F(3,51) = 1.51$, $p = .222$. Finally, there were no significant effects of any factor on subjective ratings of self-confidence.

DISCUSSION

The purpose of this study was to examine the differential impact of information and decision automation on human performance and to investigate the costs of automation unreliability in a simulated command and control task. Three different types of decision automation were examined. When automation was reliable it enhanced performance. In particular, decision automation significantly reduced the time for target engagement decisions. This is important for reducing the overall "sensor-to-shooter" time in tactical C^2 operations. However, the automation was imperfect, and when it provided an incorrect assessment the accuracy of target engagement decisions declined. When overall automation reliability was 80%, this cost of automation unreliability was greater for the three decision automation support tools than for the information automation condition. The accuracy of engagement decisions was significantly reduced, and the decision time increased, confirming the previous studies of Crocoll and Coury (1990) and Sarter and Schroeder (2001). The results demonstrate the replicability of the earlier findings in a simulated command and control task and for different forms of decision automation. The differential cost of automation unreliability for three forms of decision automation, compared to information automation, confirmed our hypothesis regarding the effects of automation unreliability. However, these effects were found only for 80%

overall automation reliability. At 60% overall automation reliability, both information and decision automation reduced performance on unreliable trials. This finding suggests that below a certain "threshold" level of reliability, automation imperfection leads to poorer performance irrespective of the type of automation. In a recent review, Wickens and Dixon (2005) suggested that human performance is sensitive to the level of automation imperfection. Our data supports this view, as we found worse performance for information automation when the overall automation reliability was 60% relative to 80%.

These findings may be interpreted via the automation model proposed by Parasuraman et al. (2000). Information automation can give the operator status information, integrate different sources of data, and may also recommend possible courses of action (Jones, et al., 2000). However, this form of automation typically does not give values to the possible courses of action, which decision automation does (Wickens, 2000). Therefore, it is possible that when the automation is highly reliable yet imperfect, performance is better with an information support tool because the user continues to generate the values for the different courses of action, and hence, is not as detrimentally influenced by inaccurate information. Additionally, a user of decision automation may no longer create or explore novel alternatives apart from those provided by the automation, thus leading to a greater performance cost when the automation is unreliable.

The effects of two levels of overall automation reliability allowed for a comparison to previous work on automation complacency in which automation reliability was varied (May, et al., 1993; Parasuraman, et al., 1993). The results showed that there was a greater cost of unreliable decision automation with high (80%) than with low (60%) overall automation reliability. This finding indicates that participants rely to a greater extent on automation when it is more reliable, even though it is imperfect. This is also consistent with a study by May et al. (1993) in which they concluded that the more reliable the automation, the more operators had reason to believe the automation to be dependable. As a result, when the automation was performing reliably, complacency increased, leading to poorer operator detection rates when it failed. The results of the present study parallel this finding, and both are examples of one of the paradoxes of automation. The more reliable the automation, the greater its detrimental effect when it fails.

One problem with this interpretation, however, is that the subjective ratings of participant trust in automation did not provide supporting evidence. While trust ratings were greater for decision than information automation, somewhat surprisingly, they did not vary with overall automation reliability as expected. Wiegmann, Rich, and Zhang (2001) also observed a disassociation between subjective ratings of automation reliability and objective performance. It was suggested that the two measures might not consistently reflect the same underlying construct of automation trust (Wiegmann, et al., 2001). The relationship between subjective ratings of trust and automation usage is complex and depends on a range of

environmental and individual factors (Dzindolet, Pierce, Beck & Dawe, 2000; see review by Lee & See, 2004). Further research on the relationship between subjective trust measures and actual automation usage would be helpful in understanding why trust ratings and automation reliance sometimes do and at other times do not dissociate.

An additional feature of the present study is that the first automation failure effect was examined and tested statistically. Based on previous research (Merlo et al., 2000; Stanton, Young, & McCaulder, 1997), it was hypothesized that participants would perform poorly the first time the automation failed, but that after some experience with unreliable automation, participants would be able to calibrate their trust accordingly; thus reducing complacency and improving performance. The results showed that performance changed after exposure to unreliable automation. In general, the reduction in decision accuracy with unreliable automation was less marked with each subsequent failure; this change was significant for both medium and high decision automation. However, we did not obtain statistically significant evidence for increased correct response rate for all failures subsequent to the first failure in all conditions. The reason could be simply due to the low power associated with analyzing data from single trials.

These results partially support the findings of Wickens and Xu (2002). Prior literature on the first failure effect is somewhat contradictory. Some studies have found little or no evidence of such effects (Davison & Wickens, 1999; Wickens, Gempler, & Morphew, 2000). One reason may be the use of displays that integrate raw data, thus supporting the operator better and reducing complacency effects (Molloy & Parasuraman, 1994; Wickens et al., 2000). Another reason for not finding first automation failure effects, as proposed by Wickens and Xu (2002), pertains to the instructions that participants receive. Prior instructions that the automation may be imperfect, as well as experience with unreliable automation in practice trials may minimize the human performance consequences of automation failure (Wickens, Helleberg & Xu, 2002). While participants in this study did not have experience with the unreliable trials during practice blocks, they were specifically instructed that the automation was highly reliable but not 100% reliable; yet, first automation failure effects were still found.

The greater detrimental impact on user performance of imperfect yet highly reliable decision automation compared to information automation may be particularly apparent in high-workload, time-stressed work environments, as in the present simulation study, and in conditions of high-tempo engagement in the battlefield. The target engagement task was relatively complex, had to be time shared with a communications task, and had to be completed within 10 s. Such conditions may have encouraged reliance on the decision choice suggested by the automation, particularly if there was insufficient time for the user to check the information sources to verify the automated advice. When the conditions permit such verification, however, the costs of imperfection in decision automation may be reduced. Lorenz, Di Nocera, Röttger, and Parasuraman

(2002) found that a high level of decision automation did not lead to poorer decision-making performance (compared to a lower or moderate level of automation) when an automated fault management system failed. They attributed this lack of a cost of imperfection in high-level decision automation to the ability of the users to query the system, inspect the raw information sources, and verify or negate the automated advice.

Limitations

The sensor to shooter micro-world used in this study may not be fully representative of command and control automated environments. Participants did not have to account for real-world stressors including fatigue, environmental factors, and fatal implications. Furthermore, the cognitive demands involved in the communications task may not have been sufficiently challenging. In the simulated environment participants simply had to click on a 'call answer' button, whereas in military environments communications involves more detail; for instance, operators may need to tune a radio to a particular frequency. The limited complexity of the sensor to shooter task may explain the lack in accuracy differences between manual and reliable automation conditions. Further, participants performed better with less reliable automation which may be a result of rejecting the automation entirely and opting for manual control. It may be difficult for operators of future automated military systems to discard automation because of the vast amount of information that would need to be gathered and processed.

Another possible limitation of the study was the decision to include perfect trials between the two varying levels of overall automation reliability. We deliberately made this choice in our experimental design despite the danger that presenting 100% reliable automation followed by either 60% or 80% reliability could have led to a possible "averaging" effect of automation reliabilities (e.g. (60 + 100)/2 vs. (80 + 100)/2 = 80% vs. 90%), so that any difference between the two unreliable levels would be potentially obscured. However, we felt it was more important to allow operators to recover in an effort to encourage automation usage and hence preceded each unreliable automation block with perfectly reliable automation. Additionally, it could be that simply providing participants with a block of reliable trials may not have been sufficient to rebuild trust in the automation. If this were the case, we would have expected to find differences between responses to the first automation failure in the experiment, and responses to first failures occurring in subsequent conditions. We found that regardless of whether the participants correctly or incorrectly responded to the initial first failure in the experiment, they still incorrectly respond to first failures in subsequent automated conditions.

Practical Implications

The current study has potentially important implications for automation use and the design of real world decision support tools. Human automation teams may break down when automation is unreliable because an operator may no longer experience benefits from the automation. Thresholds for automation trust need to be better understood specifically regarding the differential impact of information and decision automation. The present research in conjunction with previous studies (Crocoll & Coury 1990; Sarter & Schroeder, 2001) suggests that there may be a larger gap between the acceptability of information and decision automation when the system is not 100% reliable. The problem is the trend in automation tools that support the operator by completing the integration process and providing a solution with much of the raw data removed. This is problematic because it removes the operator's understanding of the information and moves the operator further away from the decision-making process. As a result, when the automation is unreliable, system performance degrades particularly when decision automation is provided. Therefore, if reliable decision automation cannot be guaranteed, it is recommended that information automation be provided to support the user or at least a low level of decision automation versus a highly autonomous decision support tool. Future research in this area may allow for the use of more sophisticated support tools to be used in dynamic environments while minimizing performance costs and improving safety.

ACKNOWLEDGEMENTS

This work was supported by Grant DAAD17-00-p-0366 from the Army Research Laboratory, Ft. Huachuca, Grant GSRP ORG 364230 from NASA Langley Research Center and by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. Many thanks to Ulla Metzger, Bernd Lorenz, Marla Zinni, and Peter Squire. Lastly, thanks to Xiong Jiang for programming the sensor to shooter simulation.

REFERENCES

Adams, T. K. (2001). Future warfare and the decline of human decision-making. *Parameters* (2001-2002, Winter), 57-71.

Barnes, M. J. (2003). *The human dimension of battlespace visualization: Research and design issues.* (Technical Report ARL-TR-2885). Aberdeen Proving Ground, MD: Army Research Laboratory.

Crocoll, W. M., & Coury, B. G. (1990). Status or recommendation: Selecting the type of information for decision aiding. *Proceedings of the Human Factors Society, 34th Annual Meeting*, Santa Monica, CA, 1524-1528.

Davison, H., & Wickens, C. D. (1999). *Rotorcraft hazard cueing: The effects on attention and trust* (Technical Report ARL-99-5/NASA-99-1). Savoy, IL: University of Illinois, Aviation Research Lab.

Department of the Army. (2003). *THAAD theatre high altitude area defense missile system. The website for defense industries - army*: Author. Retrieved November 17, 2003 from source www.army-technology.com/projects/thaad/

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2001). *A framework of automation use* (Technical Report ARL-CR-2412). Aberdeen Proving Ground, MD: Army Research Lab.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics, 42(3), 462-292.

Galster, S. M., Bolia, R. S ., & Parasuraman, R. (2002). Effects of information and decision-aiding cueing on action implementation in a visual search task. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA, 438-442.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: Elsevier Science Pub. Co.

Jones, P. M., Wilkins, D. C., Bargar, R., Sniezek, J., Asaro, P., Danner, N., & et al. (2000). CoRAVEN: Knowledge-based support for intelligent analysis. In *Federated Laboratory 4th Annual Symposium: Advanced Displays & Interactive Displays Consortium* (pp.89-94). University Park, MD: US Army Research Laboratory.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human Computer Studies, 40,* 153-184.

Lee, J. D., & Sanquist, T. F. (1996). Maritime automation. In R. Parasuraman & M. Mouloua (Eds.), *Human performance in automated systems: Recent research and trends.* (pp. 365-384). Hillsdale, NJ: Erlbaum.

Lee, J., & See, J. (2004). Trust in automation and technology: Designing for appropriate reliance. *Human Factors, 46,* 50-80.

Lorenz, B., Di Nocera, F., Röttger, S., & Parasuraman, R. (2002). Automated fault management in a simulated space flight micro-world. *Aviation, Space and Environmental Medicine, 73*, 886-897.

May, P. A., Molloy, R., & Parasuraman, R. (1993, October). *Effects of automation reliability and failure rate on monitoring performance in a multi-task environment*. Paper presented at the Human Factors and Ergonomics Society 37th Annual Meeting. Santa Monica, CA.

Merlo, J. L., Wickens, C., Yeh, M. (2000). Effect of reliability on cue effectiveness and display signaling. *Proceedings of the 4th Annual Army Federated Laboratory Symposium.* College Park, MD: Army Research Laboratory, 27-31.

Molloy, R., & Parasuraman, R. (1994). Automation-induced monitoring inefficiency: The role of display integration and redundant color coding. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Recent research and trends* (pp. 224-228). Hillsdale, NJ: Erlbaum.

Ott, R. L. (1993). *An introduction to statistical methods and data analysis*. (pp. 366-367). Belmont, CA: Wadsworth.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". *The International Journal of Aviation Psychology, 3*, 1-23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30*, 286-297.

Sarter, N., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors, 43*, 573-583.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators.* MIT, Man Machine Systems Laboratory, Cambridge, MA.

Stanton, N.A., Young, M. & McCaulder, B. (1997). Drive-by-wire: the case of mental workload and the ability of the driver to reclaim control. *Safety Science, 27* (2-3), 149-159.

Wickens, C. D. (2000). *Imperfect and unreliable automation and its implications for attention allocation, information access and situation awareness* (ARL-00-10/NASA-00-2). Moffet Field, CA: NASA Ames.

Wickens, C. D. & Dixon, S. R. (2005). *Is there a magic number 7 (to the minus 1)? The benefits of imperfect diagnostic automation: A synthesis of the literature* (Technical Report AHFD-05-01 / MAAD-05-1). Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of traffic displays in aircraft traffic avoidance. *Transportation Human Factors Journal, 2(2),* 99-126.

Wickens, C, D., Helleberg, J., & Xu, X. (2002). Pilot maneuver, choice, and workload in free flight. *Human Factors, 44*, 171-188.

Wickens, C. D. & Xu, X. (2002). *Automation trust, reliability and attention.* (Technical Report AHFD-02-14 / MAAD-02-2). Savoy, IL: University of Illinois, Aviation Research Lab.

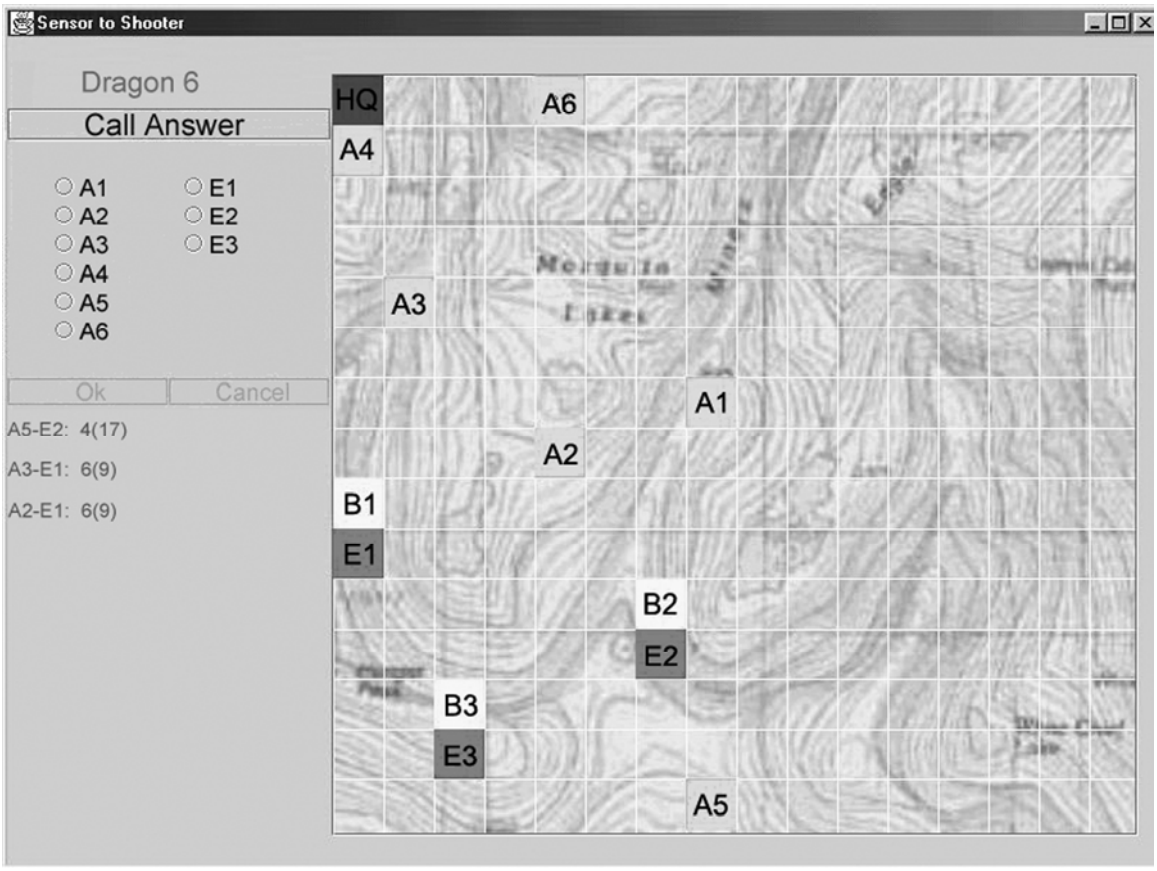
Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics, 2*, 352-367.

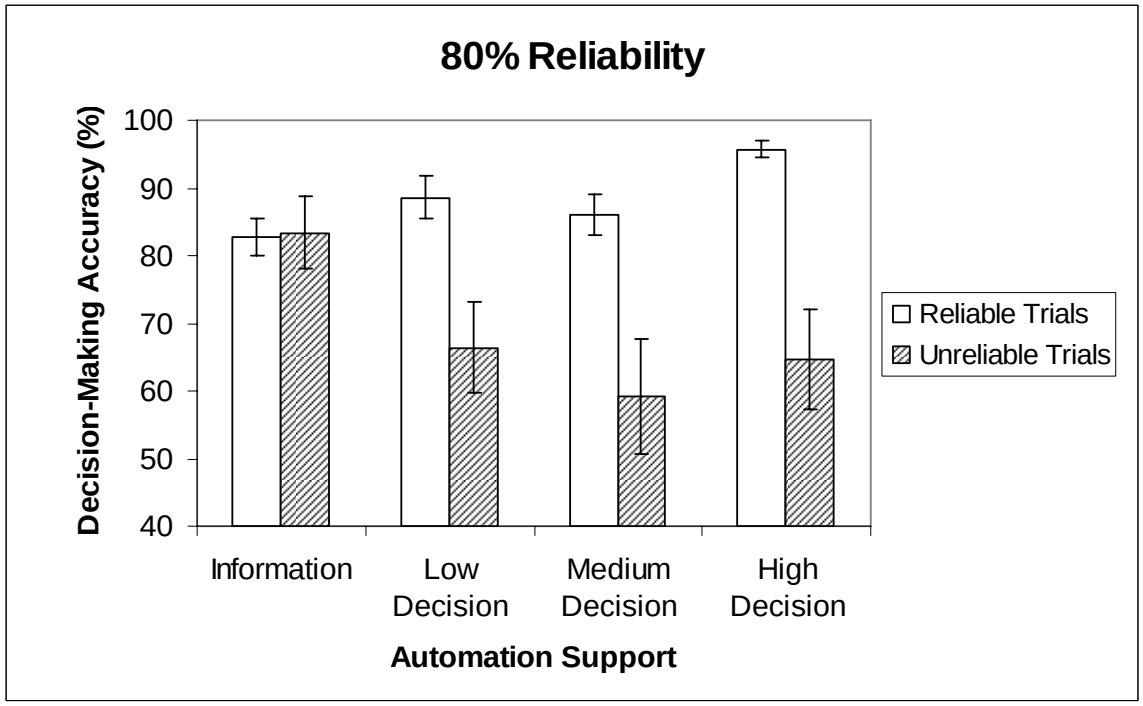
FIGURE LEGENDS

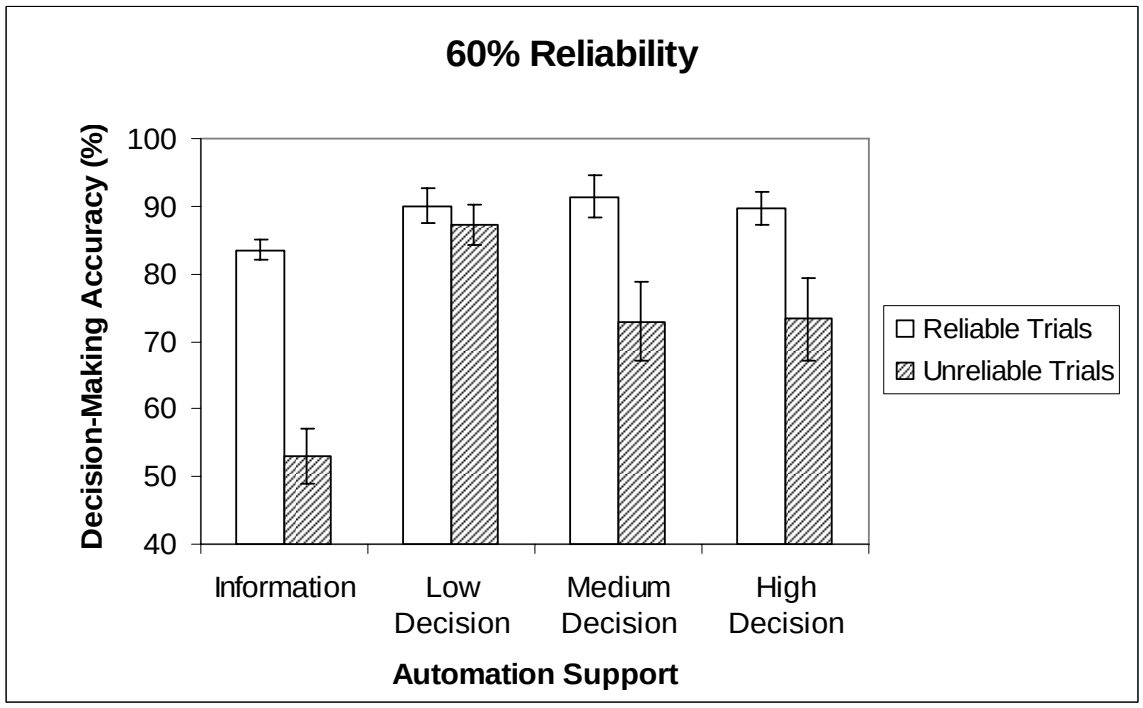
Figure 1: Sensor to shooter display with medium decision automation.

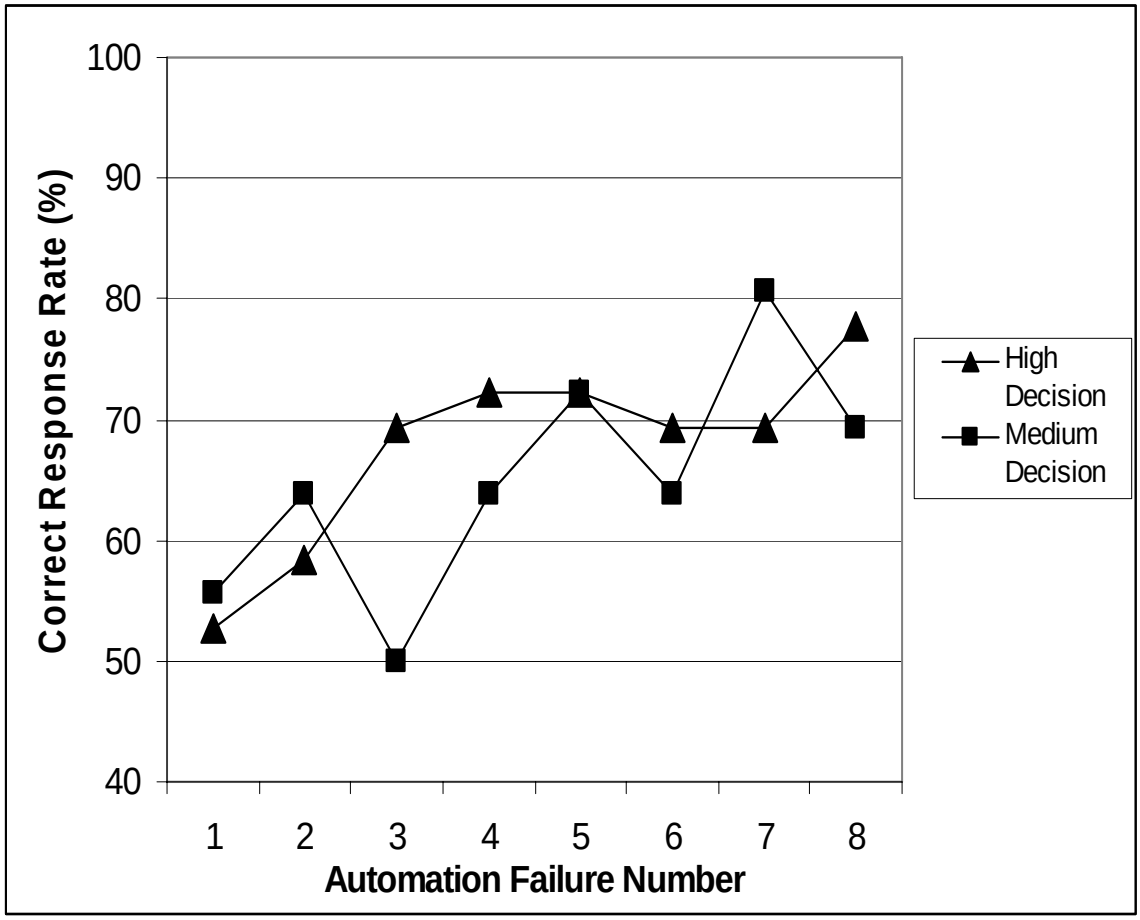
Figure 2. Decision-making accuracy for reliable and unreliable trials as a function of type of automation support and (A) 80%, (B) 60% overall automation reliability.

Figure 3: The proportion of participants that made a correct decision following an automation failure. Averaging across automation reliability conditions (60% & 80%); medium and high decision automation conditions are shown.


Sensor to Shooter
Dragon 6
Call Answer
A1
A2
A3
A4
A5
A6
E1
E2
E3
Ok
Cancel
A5-E2: 4(17)
A3-E1: 6(9)
A2-E1: 6(9)
HQ
A6
A4
A3
A1
A2
B1
E1
B2
E2
B3
E3
A5


80% Reliability
Decision-Making Accuracy (%)
100
90
80
70
60
50
40
Information
Low Decision
Medium Decision
High Decision
Automation Support
Reliable Trials
Unreliable Trials


60% Reliability
Decision-Making Accuracy (%)
100
90
80
70
60
50
40
Information
Low Decision
Medium Decision
High Decision
Automation Support
Reliable Trials
Unreliable Trials


100
90
80
70
60
50
40
Correct Response Rate (%)
1
2
3
4
5
6
7
8
Automation Failure Number
High Decision
Medium Decision

BIOGRAPHIES

Ericka Rovira is a Ph D Candidate in the Applied Experimental Psychology Program at The Catholic University of America and is currently a research affiliate at the Arch Lab, George Mason University. She received her M.A. in psychology from The Catholic University of America, USA in 2003.

Kathleen McGarry is a Ph D Candidate in the Applied Experimental Psychology Program at The Catholic University of America and is currently a research affiliate at the Arch Lab, George Mason University. She received her M.A. in psychology from The Catholic University of America, USA in 2004.

Raja Parasuraman is Professor of Psychology at the Arch Lab, George Mason University. He received his Ph.D. in psychology from Aston University, UK in 1976.

1

Raja Parasuraman and Matthew Rizzo

Introduction to Neuroergonomics

Neuroergonomics is the study of brain and behavior at work (Parasuraman, 2003). This interdisciplinary area of research and practice merges the disciplines of neuroscience and ergonomics (or human factors) in order to maximize the benefits of each. The goal is not just to study brain structure and function, which is the province of neuroscience, but also to do so in the context of human cognition and behavior at work, at home, in transportation, and in other everyday environments. Neuroergonomics focuses on investigations of the neural bases of such perceptual and cognitive functions as seeing, hearing, attending, remembering, deciding, and planning in relation to technologies and settings in the real world. Because the human brain interacts with the world via a physical body, neuroergonomics is also concerned with the neural basis of physical performance—grasping, moving, or lifting objects and one's limbs.

Whenever a new interdisciplinary venture is proposed, it is legitimate to ask whether it is necessary. To answer this query, we show how the chapters in this book, as well as other work, demonstrate that neuroergonomics provides added value, beyond that available from "traditional" neuroscience and "conventional" ergonomics, to our understanding of brain function and behavior as it occurs in the real world. The guiding principle of neuroergonomics is that examining how the brain carries out the complex tasks of everyday life—and not just the simple, artificial tasks of the research laboratory—can provide important benefits for both ergonomics research and practice. An understanding of brain function can lead to the development and refinement of theory in ergonomics, which in turn will promote new, far-reaching types of research. For example, knowledge of how the brain processes visual, auditory, and tactile information can provide important guidelines and constraints for theories of information presentation and task design. The basic premise is that the neuroergonomic approach allows the researcher to ask different questions and develop new explanatory frameworks about humans and work than an approach based solely on the measurement of the overt performance or subjective perceptions of the human operator. The added value that neuroergonomics provides is likely to be even greater for work settings such as modern semiautomated systems (Parasuraman & Riley, 1997), where measures of overt user behavior can be difficult to obtain (Kramer & Weber 2000).

Some Examples of Neuroergonomics Research

Aviation

The following examples illustrate the value of the neuroergonomic approach. Historically, the greatest influence of human factors on technological design has been in the domain of aviation, specifically in the design of displays and controls in the aircraft cockpit (Fitts, Jones, & Milton, 1950; Wiener & Nagel, 1988). With the worldwide growth in airline travel, new proposals for air traffic management have been put forward. Implementing these proposals requires new cockpit technologies. Consider a new traffic-monitoring system that is to be installed in the cockpit to portray to the pilot other aircraft that are in the immediate vicinity, showing their speed, altitude, flight path, and so on, using color-coded symbols on a computer display. Various types of neuroergonomic research, both basic and applied, can inform the design of this system. For example, designers may wish to know what features of the symbols (e.g., shape, intensity, motion, etc.) serve to best attract the pilot's attention to a potential intruder in the immediate airspace. At the same time, there may be a concern that the presentation of traffic information, while helping the pilot monitor the immediate airspace, may increase the pilot's overall mental workload, thereby degrading the performance of the primary flight task. Although subjective or performance measures could be used to evaluate this possibility, a neuroergonomic approach can provide more sensitive evaluation of any impact on flight performance. It may also lead the researcher to ask new and potentially more profitable questions about attention allocation than before. Measures of brain function that reflect visual attention and oculomotor control can help determine the impact of the new display on the pilot's visual scanning and attentional performance (see chapter 7, this volume). Finally, neuroergonomic evaluation of the manual and physical demands involved in interacting with the information panels and controls of the new traffic-monitoring system would also be required for this system to be used effectively and safely by pilots (see chapter 15, this volume).

Driving

Neuroergonomics is also relevant to assessing interactions between the eye, the brain, and the automobile (Rizzo & Kellison, 2004). Functional magnetic resonance imaging (fMRI) permits noninvasive dynamic imaging of the human brain (see chapter 4, this volume). Analytic approaches to fMRI data, such as independent component analysis, can reveal meaningful patterns in data sets collected in subjects performing complex tasks that capture elements of automobile driving. Preliminary application of these approaches suggests that multiple neural regions, including the frontoparietal, cerebellar, and occipital areas, are differentially activated by various aspects of the driving task, such as speed control. It is also possible to relate physiological correlates of impending sleep (microsleeps) derived from electroencephalographic (EEG) activity recordings of brain activity to imminent declines in driver performance (Paul, Boyle, Rizzo, & Tippin, 2005). Finally, naturalistic studies of driver behavior provide unique evidence of long-range human interactions, strategies, and tactics of "the brain in the wild" (see chapter 8, this volume).

Neuroenginering

A third example concerns the use of brain signals as an additional communication channel for human interaction with both the natural and the human-made environment. This area of research and practice, sometimes also called *neuroengineering* or *brain-computer interface* (BCI), has had significant progress in recent years. In this approach, different types of brain signals are used to control external devices without the need for motor output, which would be advantageous for individuals who either have only limited motor control or, as in the case of "locked-in" patients with amyotrophic lateral sclerosis, virtually no motor control. The idea follows naturally from the work on "biocybernetics" in the 1980s pioneered by Donchin and others (Donchin, 1980; Gomer, 1981) but has progressed beyond the earlier achievements with technical developments in recording of brain activity in real time.

BCIs allow a user to interact with the environment without engaging in any muscular activity, for

example, without the need for hand, eye, foot, or mouth movement. Instead, the user is trained to engage in a specific type of mental activity that is associated with a unique brain electrical "signature." The resulting brain potentials are recorded, processed, and classified in such a way as to provide a control signal in real time for an external device. Applications have used a variety of different measures of brain electrical activity. Invasive methods include recording of field potentials and multiunit neuronal activity from implanted electrodes; this technique has been reported to be successful in controlling robotic arms (Nicolelis, 2003). Such invasive recording techniques have superior signal-to-noise ratio but are obviously limited in use to animals or to patients with no motor functions in whom electrode implantation is clinically justified. Noninvasive BCIs have used a variety of brain signals derived from scalp EEG recordings. These include quantified EEGs from different frequency bands such as beta and mu waves (Pfurtscheller & Neuper, 2000), event-related potentials (ERPs) such as P300 (Donchin, Spence, & Wijesinghe, 2000), and contingent negative variation (Birbaumer et al., 1999). BCIs based on these signals have been used to operate voice synthesizers, control cursor movements on a computer display, and move robotic arms.

Virtual Reality

Virtual reality (VR) is particularly relevant to neuroergonomics because VR can replicate situations with greater control than is possible in the real world, allowing behavioral and neural measures of the mind and brain at work in situations that are impractical or impossible to observe in the real world. In doing so, VR can be used to study the performance of human operators engaged in hazardous tasks without putting them and others at risk for injury (see chapter 17, this volume). For example, VR can be used to study the influence of disease, drugs, fatigue, or in-vehicle technologies (such as cell phones) on aircraft piloting and automobile driving, to study how to reduce the risk of falls in the elderly, and to train students to avoid novice misjudgments and errors in performing critical medical procedures, flying aircraft, and operating heavy machinery. VR is particularly useful in workers whose jobs require spatial awareness, complex motor skills, or decisions that require evaluation of multiple possible responses amid changing contingencies, and is also proving to be useful for therapy and rehabilitation of persons with motor, cognitive, and psychiatric impairments.

Conceptual, Theoretical, and Philosophical Issues

The constituent disciplines of neuroergonomics—neuroscience and ergonomics/human factors research—are both twentieth-century, post–World War II fields. The spectacular rise of neuroscience toward the latter half of that century and the smaller but no less important growth in human factors research can both be linked to technological developments in computer science and engineering. The brain imaging technologies that have revolutionized modern neuroscience (e.g., fMRI) and the sophisticated automated systems that have stimulated much human factors work (e.g., the aircraft flight management system) were both made possible by these engineering developments. Nevertheless, the two fields have developed independently.

Traditionally, ergonomics has not paid much attention to neuroscience or to the results of studies concerning brain mechanisms underlying human perceptual, cognitive, affective, and motor processes. At the same time, neuroscience and its more recent offshoot, cognitive neuroscience, has only been recently been concerned with whether its findings bear any relation to human functioning in real (as opposed to laboratory) settings. Recent calls to move neuroscience "beyond the bench" ("Taking Neuroscience beyond the Bench," 2002) include studies of group social behavior (Cacciopo, 2002) and the development of neural prosthetics for control of robots, home automation, and other technologies for physically disabled people (see chapter 19, this volume).

The relative neglect by ergonomics of human brain function is understandable given that this discipline had its roots in a psychology of the 1940s that was firmly in the behaviorist camp. More recently, the rise of cognitive psychology in the 1960s influenced human factors, but for the most part neuroscience continued to be ignored by cognitive theorists, a state of affairs consistent with a functionalist approach to the philosophy of mind (Dennett, 1991). Such an approach implies that

the characteristics of neural structure and functioning are largely irrelevant to the development of theories of mental functioning. Cognitive psychology (and cognitive science) also went through a functionalist period in the 1970s and 1980s, mainly due to the influence of researchers from artificial intelligence and computer science. However, the recent influence of cognitive neuroscience has led to a retreat from this position. Cognitive neuroscience proposes that neural structure and function constrain and in some cases determine theories of human mental processes (Gazzaniga, 2000).

If neuroscience has freed cognitive science from rigid functionalism, then ergonomics may serve to liberate it from a disembodied existence devoid of context and provide it an anchor in the real world. Even though researchers are aware of the importance of ecological validity, modern cognitive psychology (with a few exceptions) tends to study mental processes in isolation, apart from the artifacts and technologies of the world that require the use of those processes. Technology, particularly computers, can be viewed as an extension of human cognitive capability. Related to this view is the framework of cognitive engineering, in which humans and intelligent computer systems constitute "joint cognitive systems" (Hutchins, 1995; Roth, Bennett, & Woods, 1987). Furthermore, much human behavior is situated and context dependent. Context is often defined by and even driven by technological change. How humans design, interact with, and use technology—the essence of ergonomics—should therefore also be central to cognitive science.

The idea that cognition should be considered in relation to action in the world has many antecedents. Piaget's (1952) work on cognitive development in the infant and its dependence on exploration of the environment anticipated the concept of situated or embodied cognition. Clark (1997) also examined the characteristics of an embodied mind that is shaped by and helps shape action in a physical world. If cognitive science should therefore study the mind not in isolation but in interaction with the physical world, then it is a natural second step to ask how to design artifacts in the world that best facilitate that interaction. This is the domain of ergonomics or human factors. Neuroergonomics goes one critical step further. It postulates that the human brain, which implements cognition and is itself shaped by the physical environment, must also be examined in interaction with the environment in order to understand fully the interrelationships of cognition, action, and the world of artifacts.

Currently, a coherent body of concepts and empirical evidence that constitutes neuroergonomics theory does not exist. Of course, broad theories in the human sciences are also sparse, whether in ergonomics (Hancock & Chignell, 1995) or in neuroscience (Albright, Jessell, Kandel, & Posner, 2001). Sarter & Sarter (2003) proposed that neuroergonomics must follow the same reductionist approach of cognitive neuroscience in order to develop viable theories. There are small-scale theories that could be integrated into a macrotheory, but which would still pertain only to a specific domain of human functioning. For example, neural theories of attention are becoming increasingly well specified, both at the macroscopic level of large-scale neural networks (Parasuraman, 1998; Posner, 2004) and at the level of neuronal function and gene expression (Parasuraman, Greenwood, Kumar, & Fossella, 2005; Sarter, Givens, & Bruno, 2001). At the same time, psychological theories of attention have informed human factors research and design (Wickens & Hollands, 2000). Difficult though the task may be, one can envisage amalgamation of these respective theories into a neuroergonomic theory of attention. Integration across a broader range of functional domains, however, is as yet premature.

Methods

A number of methods have been developed for use in neuroergonomic research and practice. Among these are brain imaging techniques, which have been influential in the development of the field of cognitive neuroscience. Brain imaging techniques can be roughly divided into two classes. The first group of techniques is based on measurement of cerebral hemodynamics (blood flow), such as positron emission tomography (PET), fMRI, and transcranial Doppler sonography (TCD). The second group of methods involves measurement of the electromagnetic activity of the brain, including EEG, ERPs, and magnetoencephalography (MEG). For a review of brain imaging techniques for use in studies of cognition and human performance, see Cabeza and Kingstone (2001).

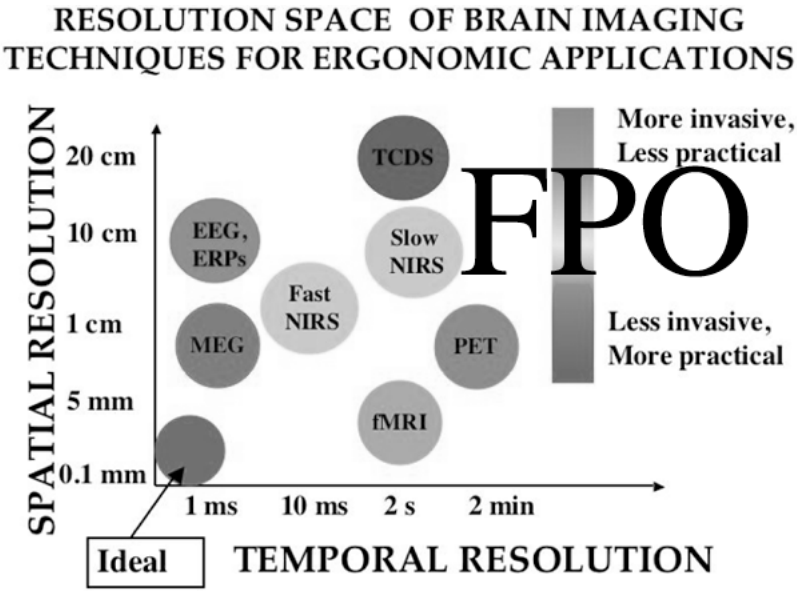


Figure 1.1. Resolution space of brain imaging techniques for ergonomic applications. Trade-offs between the criteria of the spatial resolution (*y*-axis) and temporal resolution (*x*-axis) of neuroimaging methods in measuring neuronal activity, as well as the relative noninvasiveness and ease of use of these methods in ergonomic applications (color code). EEG = electronencephalography; ERPs = event-related potentials; fMRI = functional magnetic resonance imaging; MEG = magnetoencenphalography; NIRS = near-infrared spectroscopy; PET = positron emission tomography; TCDS = transcranial doppler sonography. See also color insert.

PET and fMRI currently provide the best noninvasive imaging techniques for the evaluation and localization of neural activity. However, these methods suffer from two drawbacks. First, their temporal resolution is poor compared to electrophysiological techniques such as ERPs. Second, their use is restricted to highly controlled lab environments in which participants must not move. This limits their use for examining the neural basis of performance in more complex tasks with a view to ergonomic applications, as in flight, driving simulation, or the use of virtual reality systems, although components of complex task performance are being studied (Peres, Van de Moortele, & Pierard, 2000; Calhoun et al., 2002; see also chapter 4, this volume). Optical imaging techniques such as fast near-infrared spectroscopy (NIRS) may provide both spatial and temporal resolution and the ability to be used in neuroergonomic applications (see chapter 5, this volume).

An overview of the relative merits and disadvantages of these various techniques is shown in figure 1.1. This illustration is a variant of a representation of the spatiotemporal resolution of brain imaging methods first described by Churchland and Sejnowski (1988). The ease of ergonomic application (color code) has been added to the trade-off between the criteria of spatial resolution and temporal resolution in measuring neuronal activity. Currently, there is no one technique that reaches the ideal (blue circle) of 0.1 mm spatial resolution, 1 ms temporal resolution, and ease of use in ergonomics.

In addition to brain imaging methods, oculomotor techniques can provide additional tools for neuroergonomics researchers. With the advent of low-cost, high-speed systems for measuring different types of eye movements and increasing knowledge of the underlying neural systems, oculomotor measures can provide important information not available from traditional measurement of response accuracy and speed (see chapter 7, this volume).

It should be noted that the use of brain imaging or oculomotor measures need not be a defining characteristic of neuroergonomic research. A neuroergonomic study may use behavioral measures or a computational analysis; however, in each case the

performance measure or the computational model is related to a theory of brain function.

Consider the following example. Suppose that as a result of the manipulation of some factor, performance on a target discrimination task (e.g., detection of an intruder aircraft in the cockpit traffic-monitoring example discussed previously) in which location cues are provided prior to the target yields the following results: reaction time (RT) to the target when preceded by an invalid location cue is disproportionately increased, while that to a valid cue is not. This might happen, for example, if the cue is derived from the output of an automated detection system that is not perfectly reliable (Hitchcock et al., 2003). In simple laboratory tasks using such a cueing procedure, there is good evidence linking this performance pattern to a basic attentional operation and to activation of a specific distributed network of cortical and subcortical regions on the basis of previous research using noninvasive brain imaging in humans, invasive recordings in animals, and performance data from individuals who have suffered damage to these brain regions (Posner & Dehaene, 1994). One could then conduct a study using the same cueing procedure and performance measures as a behavioral assay of the activation of the neural network in relation to performance of a more complex task in which the same basic cognitive operation is used. If the characteristic performance pattern was observed—a disproportionate increase in RT following an invalid location cue, with a normal decrease in RT following a valid cue—then one could argue that the distributed cortical/subcortical network of brain regions is likely to have been involved in task performance. This would then enable the researcher to link the full body of neuroscience work on this aspect of attentional function to performance on the complex intruder-detection task. Thus, even though no physiological index was used, and although the same performance measure (RT) was used as in a traditional ergonomic analysis, the type of question asked and the explanatory framework can be quite different in the neuroergonomic approach.

Finally, a neuroergonomic study could also involve a computational analysis of brain or cognitive function underlying performance of a complex task. So long as the analysis was theoretically driven and linked to brain function, the study would qualify as neuroergonomic even though no physiological index was used. Several computational models of human performance have been developed for use in human factors (Pew & Mavor, 1998). Of these, models that can be linked, in principle, to brain function—such as neural network (connectionist) models (O'Reilly & Munakata, 2000)—would be of relevance to neuroergonomics.

Neuroergonomics and Neuropsychology

Neuropsychology and related fields (e.g., behavioral neurology, clinical and health psychology, neuropsychiatry, and neurorehabilitation) have also helped pave the way for neuroergonomics. Hebb (1949) used the term *neuropsychology* in his classic book *The Organization of Behavior: A Neuropsychological Theory.* The field broadly aims to understand how brain structure and function are related to specific psychological processes. The neuropsychological approach uses statistical techniques for standardizing psychological tests and scales to provide clinical diagnostic and assessment tools in normal and impaired individuals (de Oliveira Souza, Moll, & Eslinger, 2004).

Like neuroergonomics, neuropsychology is dedicated to a psychometric approach, holding that human behavior can be quantified with objective tests of verbal and nonverbal behavior, including neural states, and that these data reflect a person's states of mind and information processing. These processes can be divided into different domains, such as perception, attention, memory, language, executive functions (decision making and implementation), and motor abilities, and they can be assessed using a wide variety of techniques.

Both neuropsychology and neuroergonomics rely on principles of reliability (how repeatable a behavioral measure is) and validity (what a measure really shows about human brain and behavior). Neuropsychology has traditionally relied on paper-and-pencil tests, many of which are standardized and well understood (e.g., Lezak, 1995). The neuroergonomics approach is more rooted in technology, as indicated in this book. Novel techniques and tests are developing at a rapid pace, and guidelines and standards are going to be needed.

Contributions to Neuroergonomics from Other Fields: Genetics, Biotechnology, and Nanotechnology

While we have emphasized the contribution of neuroscience to neuroergonomics in this chapter, developments in other fields are also affecting the study of human brain function at work. Three such fields are molecular genetics, biotechnology, and nanotechnology, and we briefly consider their relevance here.

As discussed previously, cognitive psychology has increasingly capitalized on findings from neuroscience. More recently, the study of individual differences in cognitive function is being influenced by developments in molecular genetics and, in particular, the impressive results of the Human Genome Project. Much of what we know about the genetics of cognition has come from twin studies in which identical and fraternal twins are compared to assess the heritability of a trait. This paradigm has been widely used in behavioral genetics research for over a century. For example, the method has been used to show that general intelligence, or *g*, is highly heritable (Plomin & Crabbe, 2000). However, this approach cannot identify the particular genes involved in intelligence or the cognitive components of *g*. Recent advances in molecular genetics now allow a different, complementary approach to behavioral genetics, called *allelic association*. This method has been applied to the study of individual differences in cognition in healthy individuals, revealing evidence of modulation of cognitive task performance by specific neurotransmitter genes (Fan, Fossella, Sommer, Wu, & Posner, 2003; Greenwood, Sunderland, Friz, & Parasuraman, 2000; Parasuraman et al., 2005). This work is likely to provide the basis not only for improved understanding of the neural basis of cognition, but also for better characterization of individual differences in cognition. That, in turn, can have an impact on important human factors issues such as selection and training.

Reliable quantification of individual differences in cognitive function will have obvious implications for selection of operators for occupations that demand a high workload. While it would be premature to state that the molecular genetic approach to cognition has immediate applications to selection, further programmatic research on more complex cognitive tasks will undoubtedly lead to progress in such an endeavor. The postgenomic era has clearly demonstrated that inheritance of a particular genotype only sets a range for the phenotypic expression of that genotype, with the exact point within that range being determined by other genetic and environmental factors. Genomic analysis allows for a much more precise specification of that range for any phenotype, and for linking phenotypic variation to specific genetic polymorphisms. Selection and training have traditionally been considered together in human factors research and practice (e.g., Sanders & McCormick, 1983) but rarely in terms of a common biological framework. Examining the effects of normal genetic variation and of various training regimens on brain function may provide such a common framework.

The goal of neuroergonomics is to better understand the brain's functional structures and activities in relation to work and technology. In addition to molecular genetics, biotechnology can contribute to this effort by providing a means to study neuronal activities down to the molecular level. Biomimetic studies also allow for precise modeling of the human brain's activities. If the validity of such models can be established in the near future, then researchers could examine various manipulations of brain function that are not ethically possible with human participants.

The currently available measures of brain function are limited by sensor size and the inability to monitor brain function and influence function simultaneously. Nanotechnology provides the measurement tools that can achieve such dual-purpose needs. It can also provide new sensors for monitoring changes in neuronal function in otherwise undetectable brain structures. In addition, nanotechnology has the appropriate scale of operations necessary to deliver chemicals needed to precisely monitor and modify effects of neurotransmitters or encourage targeted neurogenesis, with the objective of improving human performance in certain work environments.

Although there are few current examples of the influence of biotechnology and nanotechnology on neuroergonomics, these fields are likely to have greater impact in the near future. De Pontbriand (2005) provided a cogent discussion of the potential benefits that biotechnology and nanotechnology can bring to neuroergonomics.

Overview of *Neuroergonomics: The Brain at Work*

This book represents a collective examination of the major theoretical, empirical, and practical issues raised by neuroergonomics. In this opening chapter, which forms part I, we have provided an overview of the field, covering theoretical and conceptual issues involved in the merging of cognitive neuroscience and human factors research. We have also briefly described neuroergonomic methods, but these are covered in more detail in part II, which consists of seven chapters describing different cognitive neuroscience methods: fMRI, EEG, ERPs, NIRS, TCD, and oculomotor measures. In addition, measures to track behavior and brain function in naturalistic environments are also described. Each chapter outlines the major features of each method, describes its principal merits and limitations, and gives illustrative examples of its use to address issues in neuroergonomics. We understand that readers will bring a variety of technical backgrounds to the examination of these methodological issues. Accordingly, key readings provided at the end of each chapter provide additional background for understanding some of the more technical details of each method, as needed.

Part III examines basic research in a number of different domains of cognition that have particular relevance for the understanding of human performance at work. We did not attempt to be comprehensive. Rather, we chose areas of cognition in which significant progress has been made in identifying the underlying neural mechanisms, thereby allowing for theory-driven application to human factors issues. The cognitive domains discussed are spatial cognition, vigilance, executive functions, and emotion and decision making. In addition, working memory, planning, and prospective memory are variously described in some of these chapters as well as in other sections of the book.

As the study of the brain at work, neuroergonomics must also examine the work environment. It is an undeniable fact that many work settings are stressful, induce fatigue, and are poorly designed in terms of workspace layout. Accordingly, part IV examines issues of stress, sleep loss, and fatigue, as well as the effects of the physical work environment.

Part V consists of four chapters that discuss several different domains of application of neuroergonomics. Again, we did not attempt to cover all of the application areas that are emerging as a result of the use of neuroergonomic research. We chose four: adaptive automation, virtual reality, robotics, and neuroengineering.

Neuroengineering applications are designed in part to help individuals with different disabilities that make it difficult for them to communicate effectively with the world. This area of work is covered in more detail in part VI. Four chapters describe neuroergonomic technologies that can be used to help the paralyzed, individuals with low or no vision, and those who require prostheses. A final chapter in this section is concerned with the evaluation of medical safety in health care settings.

Finally, in part VII, we close the volume by surveying prospects for the future of neuroergonomics.

Conclusion

Neuroergonomics represents a deliberate merger of neuroscience and ergonomics with the goal of advancing understanding of brain function underlying human performance of complex, real-world tasks. A second major goal is to use existing and emerging knowledge of human performance and brain function to design technologies and work environments for safer and more efficient operation. More progress has been made on the first goal than on the second, but both neuroergonomic research and practice should flourish in the future, as the value of the approach is appreciated. The basic enterprise of ergonomics—how humans design, interact with and use technology—can be considerably enriched if we also consider the human brain that makes such activities possible.

Main Points

1. Neuroergonomics is the study of brain and behavior at work.
2. Neuroergonomics attempts to go beyond its constituent disciplines of neuroscience and ergonomics by examining brain function and cognitive processes not in isolation but in relation to the technologies and artifacts of everyday life.
3. Some examples of neuroergonomics include research in the areas of aviation, driving,

brain-computer interfaces, and virtual reality.
4. Neuroergonomics is inconsistent with a purely functional philosophy of mind, in which brain structure and function are deemed irrelevant. In addition, neuroergonomics views brain and mind as influenced by context and technology.
5. Neuroergonomic methods include behavioral and performance studies, brain imaging, oculomotor measures, and computational techniques. These methods have different relative merits and disadvantages.

Key Readings

Cabeza, R. M., & Kingstone, A. (2001). *Handbook of functional neuroimaging of cognition.* Cambridge, MA: MIT Press.

Kramer, A. F., & Weber, T. (2000). Applications of psychophysiology to human factors. In J. T. Cacciopppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 794–814). New York: Cambridge University Press.

Mussa-Ivaldi, F. A., & Miller, L. E. (2003). Brain-machine interfaces: Computational demands and clinical needs meet basic neuroscience. *Trends in Cognitive Sciences, 26,* 329–334.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science, 4,* 5–20.

References

Albright, T. D., Jessell, T. M., Kandel, E. R., & Posner, M. I. (2001). Progress in the neural sciences in the century after Cajal (and the mysteries that remain). *Annals of the New York Academy of Sciences, 929,* 11–40.

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kubler, A., et al. (1999). A spelling device for the paralysed. *Nature, 398,* 297–298.

Cabeza, R., & Kingstone, A. (2001). *Handbook of functional neuroimaging of cognition.* Cambridge, MA: MIT Press.

Cacioppo, J. T. (Ed.). (2002). *Foundations in social neuroscience.* Cambridge, MA: MIT Press.

Calhoun, V. D., Pekar, J. D., McGinty, V. B., Adali, T., Watson, T. D., & Pearlson, G. D. (2002). Different activation dynamics in multiple neural systems during simulated driving. *Human Brain Mapping, 16,* 158–167.

Churchland, P. S., & Sejnowski, T. J. (1988). Perspectives in cognitive neuroscience. *Science, 242,* 741–745.

Clark, A. (1997). *Being there: Putting brain, body, and the world together again.* Cambridge, MA: MIT Press.

Dennett, D. (1991). *Consciousnesss explained.* Cambridge, MA: MIT Press.

de Oliveira Souza, R., Moll, J., & Eslinger, P. J. (2004). Neuropsychological assessment. In M. Rizzo & P. J. Eslinger (Eds.), *Principles and practice of behavioral neurology and neuropsychology.* (pp. 343–367). New York: Saunders.

de Pontbriand, R. (2005). Neuro-ergonomics support from bio- and nano-technologies. (p. 2512). In *Proceedings of the 11th International Conference on Human Computer Interaction*. Las Vegas, NV: HCI International.

Donchin, E. (1980). Event-related potentials: Inferring cognitive activity in operational settings. In F. E. Gomer (Ed.), *Biocybernetic applications for military systems* (pp. 35–42) (Technical Report MDC EB1911). Long Beach, CA: McDonnell Douglas.

Donchin, E., Spencer, K. M., & Wijesinghe, R. (2000). The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering, 8,* 174–179.

Fan, J., Fossella, J. A., Sommer, T., Wu, Y., & Posner, M. I. (2003). Mapping the genetic variation of attention onto brain activity. *Proceedings of the National Academy of Sciences USA, 100*(12), 7406–7411.

Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review, 9*, 24–29.

Gazzaniga, M. S. (2000). *The cognitive neurosciences.* Cambridge, MA: MIT Presss.

Gomer, F. (1981). Physiological systems and the concept of adaptive systems. In J. Moraal & K. F. Kraiss (Eds.), Manned systems design (pp. 257–263). New York: Plenum Press.

Greenwood, P. M., Sunderland, T., Friz, J. L., & Parasuraman, R. (2000). Genetics and visual attention: Selective deficits in healthy adult carriers of the varepsilon 4 allele of the apolipoprotein E gene. *Proceedings of the National Academy of Sciences USA, 97*(21), 11661–11666.

Hancock, P. A., & Chignell, M. H. (1995). On human factors. In J. Flach, P. Hancock, J. Caird, & K. Vicente (Eds.), *Global perspectives on the ecology of human-machine systems* (pp. 14–53). Mahwah, NJ: Erlbaum.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory.* New York: Wiley.

Hitchcock, E. M., Warm, J. S., Matthews, G., Dember, W. N., Shear, P. K., Tripp, L. D., et al. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science, 4*, 89–112.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Kramer, A. F., & Weber, T. (2000). Applications of psychophysiology to human factors. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 794–814). New York: Cambridge University Press.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford.

Nicolelis, M. A. (2003). Brain-machine interfaces to restore motor function and probe neural circuits. *Nature Reviews Neuroscience, 4*, 417–422.

O'Reilly, R. C., & Munukata, Y. (2000). *Computational explorations in cognitive neuroscience.* Cambridge, MA: MIT Press.

Parasuraman, R. (1998). The attentive brain: Issues and prospects. In R. Parasuraman (Ed.) *The attentive brain* (pp. 3–15). Cambridge, MA: MIT Press.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science, 4,* 5–20.

Parasuraman, R., Greenwood, P. M., Kumar, R., & Fossella, J. (2005). Beyond heritability: Neurotransmitter genes differentially modulate visuospatial attention and working memory. *Psychological Science*, *16,* 200–207.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230–253.

Paul, A., Boyle, L., Rizzo, M., & Tippin, J. (2005). Variability of driving performance during microsleeps. In L. Boyle, J. D. Lee, D. V. McGehee, M. Raby, & M. Rizzo (Eds.) *Proceedings of Driving Assessment 2005: The Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. (pp. 18–24). Iowa City: University of Iowa.

Peres, M., Van De Moortele, P. F., & Pierard, C. (2000). Functional magnetic resonance imaging of mental strategy in a simulated aviation performance task. *Aviation, Space and Environmental Medicine, 71,* 1218–1231.

Pew, R., & Mavor, A. (1998). *Modeling human and organizational behavior.* Washington DC: National Academy Press.

Pfurtuscheller, G., & Neuper. (2001). Motor imagery and direct brain-computer communication. *Proceedings of the IEEE, 89,* 1123–1134.

Piaget, J. (1952). *The origins of intelligence in children*. New York: Longman.

Plomin, R., & Crabbe, J. (2000). DNA. *Psychological Bulletin, 126,* 806–828.

Posner, M. I. (2004). *Cognitive neuroscience of attention.* New York: Guilford.

Posner, M. I., & Deheane, S. (1994). Attentional networks. *Trends in Neuroscience, 17,* 75–79.

Rizzo, M., & Kellison, I. L. (2004). Eyes, brains, and autos. *Archives of Ophthalmology, 122,* 641–647.

Roth, E. M., Bennett, K. B., & Woods, D. D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man-Machine Studies, 27*, 479–525.

Sanders, M. S., & McCormick, E. F. (1983). *Human factors in engineering and design.* New York: McGraw-Hill.

Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: Where top-down meets bottom-up. *Brain Research Reviews, 35*, 146–160.

Sarter, N., & Sarter, M. (2003). Neuroergonomics: Opportunities and challenges of merging cognitive neuroscience with cognitive ergonomics. *Theoretical Issues in Ergonomics Science, 4*, 142–150.

Taking neuroscience beyond the bench [Editorial]. (2002). *Nature Neuroscience*, *5*(Suppl.), 1019.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance.* New York: Longman.

Wiener, E. L., & Nagel, D. C. (1988). *Human factors in aviation*. San Diego, CA: Academic Press.

24

Matthew Rizzo and Raja Parasuraman

Future Prospects for Neuroergonomics

The preceding chapters present strong evidence for the growth and development of neuroergonomics since its inception a few years ago (Parasuraman, 2003). The ever-increasing understanding of the brain and behavior at work in the real world, the development of theoretical underpinnings, and the relentless spread of facilitative technology in the West and abroad are inexorably broadening the substrates for this interdisciplinary area of research and practice. Neuroergonomics blends neuroscience and ergonomics to the mutual benefit of both fields and extends the study of brain structure and function beyond the contrived laboratory settings often used in neuropsychological, psychophysical, cognitive science, and other neuroscience-related fields.

Neuroergonomics is providing rich observations of the brain and behavior at work, at home, in transportation, and in other everyday environments in human operators who see, hear, feel, attend, remember, decide, plan, act, move, or manipulate objects among other people and technology in diverse, real-world settings. The neuroergonomics approach is allowing researchers to ask different questions and develop new explanatory frameworks about humans at work in the real world and in relation to modern automated systems and machines, drawing from principles of neuropsychology, psychophysics, neurophysiology, and anatomy at neuronal and systems levels.

Better understanding of brain function, as outlined in the chapters on perception, cognition, and emotion, is leading to the development and refinement of theory in neuroergonomics, which in turn is promoting new insights, hypotheses, and research. For example, research on how the brain processes visual, auditory, and tactile information is providing important guidelines and constraints for theories of information presentation and task design, optimization of alerting and warning signals, development of neural prostheses, mitigation of errors by operators whose physiological profiles indicate poor functioning or fatigue, and development of robots that emulate or are part of human beings.

Specific challenges for this new field, both in the near term and beyond, are outlined in each section of this book (part II, Methods; part III, Perception and Cognition; part IV, Stress, Fatigue, and Physical Work; part V, Technology Applications; and part VI, Special Populations). In this closing chapter, we briefly discuss prospects for the future, both in the near term and the longer term. We also address some of the general challenges facing neuroergonomics research and practice. These include issues of privacy and ethics and the development of

standards and guidelines to ensure the quality and safety of a host of new procedures and applications.

The Near Future

Simulation and Virtual Reality

An imminent challenge in neuroergonomics will be to disseminate and advance new methods for measuring human performance and physiology in natural and naturalistic settings. It is likely that functional magnetic resonance imaging (fMRI) methods will be further applied to study brain activity in tasks that simulate instrumental activities of daily living, within the constraints of the scanner environment. The ability to image brain activity during complex, dynamic behavior such as driving and navigation tasks will enhance our understanding of the neural correlates of complex behaviors and the performances of neurologically normal and impaired individuals in the real world. Further use of fMRI paradigms involving naturalistic behaviors will depend on improved software design and analytic approaches such as independent component analysis to decompose and understand the complex data sets collected from people engaged in these complex tasks. This exciting future also depends on advances in brain imaging hardware, fMRI experimental design and data analyses, ever-better virtual reality (VR) environments, and stronger evidence to determine the extent to which tasks in these environments are valid surrogates for tasks in the real world.

Multidisciplinary vision and collaboration have established VR as a powerful tool to advance neuroergonomics. VR applications—using systems that vary from surrealistic, PC-based gaming platforms to high-fidelity, motion-based simulation and fully immersive, head-mounted VR—are providing computer-based synthetic environments to train, treat, and augment human behavior and cognitive performance in renditions of the real world. An unanswered question regarding the fidelity of usable VR environments is, "How low can you go?" A range of VR systems will continue to be used to investigate the relationships between perception, action, and consciousness, including how we become immersed in an environment and engaged by a task, and why we believe we are where we think we are.

Augmented reality (AR) systems combine real and artificial stimuli, often by overlaying computer-generated graphics on a stream of video images, so that the virtual objects appear to be embedded in the real world. The augmentation can highlight important objects or regions and superimpose informative annotations to help operators accomplish difficult tasks. These systems may help aircraft pilots maintain situational awareness of weather, air traffic, aircraft state, and tactical operations by using a head-mounted display that displays such information and enhances occluded features (such as the horizon or runway markings); other systems may allow a surgeon to visualize internal organs through a patient's skin. Immediate challenges in creating effective AR systems include modeling the virtual objects to be embedded in the image, precisely registering the real and virtual coordinate systems, generating images quickly enough to avoid any disconcerting lag when there is relative movement, and building portable devices that do not encumber the wearer.

Physiological Monitoring

Advances in physiological measurements will permit additional observations of brain function in simulated and real-world tasks. Electroencephalogram (EEG) data should provide further data about changes in regional functional brain systems activated by ongoing task performance to complement the evidence from fMRI, positron emission tomography (PET), and other techniques. As discussed by Gevins and Smith (chapter 2), such studies are possible because EEG patterns change predictably with changes in task load, mental effort, arousal, and fatigue, and can be assessed using algorithms that combine parameters of the EEG power spectra into multivariate functions. Although these EEG data lack the 3-D spatial resolution of brain fMRI or PET, EEG is more readily applied in tasks that resemble those that an individual might encounter in a real-world environment.

Event-related potentials (ERPs) provide additional insights and applications for neuroergonomics and are computed by averaging EEG epochs time-locked to sensory, motor, and cognitive events. Although ERPs also have lower spatial resolution than fMRI, ERP resolution is improving because of new source localization techniques. Moreover, ERPs

have better temporal resolution for evaluating neural activity than other neuroimaging techniques do. As Fu and Parasuraman (chapter 3) discuss, ERP components such as P300, N1, P1, ERN, and LRP will provide additional information relevant to neuroergonomic assessments of mental workload, attention resource allocation, dual-task performance, error detection and prediction, and motor control. Advances in device miniaturization and portability will continue to enhance the utility of EEG and ERP systems. Development of automated systems, in which human operators monitor control system functions, will provide further opportunities to use ERP-based neuroergonomic measures and theories relevant to brain function at work.

Noninvasive optical imaging tools, such as near-infrared spectroscopy (NIRS), assess neuronal activity that occurs directly after stimulus presentation or in preparation for responses and hemodynamic changes that occur a few seconds later, and these add to the neuroergonomics toolbox (which includes fMRI, PET, EEG, and ERP). As Gratton and Fabiani (chapter 5) discuss, optical imaging provides a good combination of spatial and temporal resolution of brain activity, can be deployed in a range of environments and experimental conditions, and costs little compared to magnetoencephalography, PET, and fMRI. Because optical imaging systems are relatively portable, they may be applied more commonly to map the time course of brain function and hemodynamic responses. Further advances will need to overcome the relatively low signal-to-noise ratio that affects the faster neuronal signal and the limited penetration by the technique to within a few centimeters beneath the scalp, which precludes measurements of activity in deeper brain structures (subcortical and brain stem) in adults.

Images of the brain at work will be complemented by transcranial Doppler sonography (TCD), which allows fast and mobile assessment of task-related brain activation and related effects of workload, vigilance, and sustained attention. Tripp and Warm (chapter 6) show that, like NIRS-based measurement of blood oxygenation, TCD offers a noninvasive and relatively inexpensive way to "monitor the monitor." TCD may prove especially useful to assess task difficulty and task engagement and to determine when human operators need rest and whether they may benefit from adaptive automation systems designed to flexibly allocate tasks between the operators and computer systems to mitigate operator workload and fatigue.

Eyelid closure measurements can be used to monitor fatigue and falling asleep on the job or at the wheel. McCarley and Kramer (chapter 7) show that eye movement assessments provide an additional window on perception, cognition, and how we search for meaningful information in displays and real-world scenes. Advances in gaze-contingent control procedures will enable neuroergonomics researchers to infer capability, strategies, and perhaps even the intent of operators who are inspecting the panorama in a variety of complex simulated and real-world environments. Real-time measures of where a user is gazing are already being used to develop attentive user interfaces, namely devices and computers that know where a person is looking and therefore do or do not interrupt the user accordingly (Vertegaal, 2002). Evidence from high-speed measurements of eye movements may also be combined with other measures (such as EEG, ERP optical imaging, heart rate, NIRS) to enhance the design of display devices aboard aircraft, automobiles, and industrial systems and improve real-time workload and performance assessment algorithms.

It is important to recognize what can and cannot be achieved specifically by physiological monitoring in extralaboratory environments. We must also be careful not to overstate what neuroergonomics can achieve given the noisiness of most real-world settings. Some recent programmatic efforts have set very ambitious goals for real-time assessment of operator cognitive state using physiological measures (St. John, Kobus, Morrison, & Schmorrow, 2004). Meeting these goals will require substantial effort focused on the elimination of potential artifacts that may mask the signal of interest, or, more seriously, lead to flawed assessments of operator state. Given the diversity and magnitude of artifacts that are likely to be encountered in natural settings, obtained artifact-free recordings, particularly in real time, will pose a considerable technical challenge. Nevertheless, there have been a number of promising developments in the design of miniaturized recording equipment that can withstand the rigors of operational environments. Automated artifact rejection procedures have

also been developed, which, if proven robust, would help considerably in meeting the real-time monitoring challenge (see Gevins et al., 1995; see also chapter 2, this volume).

The Brain in the Wild

In addition to physiological monitoring, neuroergonomic evaluations can also involve assessment of other aspects of human behavior in natural environments, what Rizzo and colleagues call the brain in the wild (chapter 8). New technologies are allowing the development of various implementations of "people trackers" using combinations of accelerometers, global positioning systems, video, and other sensors (e.g., to measure gross and fine body movement patterns, skin temperature, eye movements, heart rate, etc.) to make naturalistic observations of human movement and behavior in the wild. As Rizzo and colleagues discuss, these trackers can advance the goal of examining performance, strategies, tactics, interactions, and errors in humans engaged in real-world tasks, drawing from established principles of ethology. Besides issues of device development and sensor choice and placement for various classes of devices (outside looking inside, inside looking inside, inside looking outside), taxonomies are needed for classifying likely behavior from sensor output. Different sensor array implementations will provide unique data on how the brain interacts with diverse environments and systems, at work and at play, and in health, disease, or fatigue states.

Such measurement techniques are likely to be important in the assessment of stress and fatigue at work. Fatigue on the job is common in our 24-hour-service society. Operational demands in round-the-clock industries such as health care and transportation and industries that require shift work inevitably cause fatigue from sleep loss and circadian displacement, which contributes to increased cognitive errors and risk of adverse events (such as medical errors and airplane crashes). Predicting and mitigating the risks posed by physiologically based variations in sleepiness and alertness at work are key functions of neuroergonomics, as discussed in chapter 14. Near-term advances will depend on more unobtrusive tools for detecting fatigue and a better understanding of the effects of sleep-wake cycles and circadian biology on human performance.

Cross-Fertilization of Fields

Evidence from many of the above-mentioned neuroergonomic areas will continue to converge with evidence from other established fields such as animal physiology and human neuropsychology. De Pontbriand (2005) also envisaged cross-fertilization between neuroergonomics and rapidly growing fields such as biotechnology. Neuroergonomics will also continue to provide a unique source of evidence in its own right. For example, Maguire (chapter 9) explains how neuroergonomics can lead to a better understanding of plasticity and dynamics within the brain's navigation systems; she foresees an increasingly fruitful exchange whereby technological and environmental improvements will be driven by an informed understanding of how the brain finds its way in the real world.

In this spirit of interdisciplinary cross-fertilization, Grafman (chapter 11) outlines a representational framework for understanding executive functions that underpin work-related performance in the real world. This performance depends on key structures in human prefrontal cortex that mediate decision making and implementation, planning and judgment, social cognition, tactics, strategy, foresight, goal achievement, and risk evaluation. Further, Bechara recognizes that real-world decisions are critically affected by emotions, in a process that reconciles cold cognition in the cortex with processes in the more primitive limbic brain. Emotion systems are a key factor in interaction between environmental conditions and human cognitive processes at work and elsewhere in the real world. These systems provide implicit or explicit signals that recruit cognitive processes that are adaptive and advantageous for survival. Understanding the neural underpinnings and regulation of emotions and feelings is crucial for many aspects of human behavior and their disorders, including performance at work, and can help provide a model for the design of new generations of computers that interact with humans, as in chapter 18.

The Longer-Term Future

Breazeal and Picard (chapter 18) explain how findings in neuroscience, cognitive science, and human behavior inspire the development of robots with

social-emotional skills. Currently in an early stage of development, these relational robots hold promise in future applications in work productivity and education, where the human user may perform better in cooperation with the robot. For example, embedding affective technologies in learning interactions with automated systems (robotic learning companions) should reveal what affective states are most conducive to learning and how they vary with teaching styles, and this information will hone the robot's ability to learn from a person. Thus, humans will learn from machines and vice versa, parallel to the vision of Grafman (chapter 11) in which cognitive neuroscience applications will inform and improve training and evaluation methods currently employed by human factors experts.

Before they augmented our minds, machines amplified our muscles. Machines began relieving stress on human muscles as soon as humans discovered wedges, sledges, wheels, fulcrums, and pulleys. Hancock and Szalma (chapter 13) point out that machines will continue to replace human muscle power to minimize stress and fatigue as much as possible. In this futuristic vision of neuroergonomics, human intentions, indexed by interpretable brain states, will connect directly to these machines to bring intention to action. The effector might be a body part-sized electromechanical prosthesis such as an artificial limb or hand, or a substantial machine such as an exoskeleton robot, easily capable of bone-crushing labor. Additional neuroergonomic tools will monitor operators, identify signs of impending performance failure, and adjust system behavior and workload to mitigate stress.

Neuroergonomic countermeasures to stress, fatigue, and sleepiness can include systems for adaptive automation in which the user and the system can initiate changes in the level of automation, triggered in part by psychophysiological measures. Operators may come to think of these adaptive systems more as coworkers (rather than tools, machines, or computer programs) and even expect them to behave like humans. To design these adaptive systems, developers will need better information about human-machine task sharing, methods for communicating goals and intentions, and assessment of operator states of mind, including trust of and annoyance with robotic systems. These adaptive automation systems will be advantageously applied in settings where safety concerns surround the operator, system, and recipient of system services. Other potential applications might include a personal assistant, butler, secretary, or receptionist; an adaptive house; and systems aimed at training and skill development, rehabilitative therapy, and entertainment.

As discussed by Scerbo, adaptive automation systems are challenged when the wishes of two or more of the operators have conflicting goals. Such conflicts may arise in health care and can be examined at several levels using a cognitive work analysis framework in research aimed at improving medical quality and reducing errors. Addressing such complexities will depend on a better understanding of how multiple operators and their brains behave in social situations in differently organized work environments.

Neural engineering is a dimension of neuroergonomics related to the establishment of direct interactions between the nervous system and artificial devices. In this arena, technology is applied to improve biological function while lessons from biology inform technology, with cross-fertilization between molecular biology, electrophysiology, mathematics, signal processing, physics, and psychology. Brain-machine and brain-computer interfaces (BCIs) can interact with telecommunication technologies for controlling remote devices or for transmitting and receiving information across large distances, enabling operations at a distance.

As Mussa-Ivaldi and colleagues (chapter 19) discuss, advances in neuroengineering will depend on better knowledge and models of brain functions, at all levels ranging from synapses to systems, and on learning how to represent these models in terms of discrete or continuous variables. In this vein, Poggel and colleagues (chapter 21) address how the brain encodes visual information to learn how to "talk" to the brain and restore vision in patients with visual impairment due to retinal lesions. Meeting this goal will depend on a better understanding of plasticity, perception, low-level signal processing, and top-down influences such as attention.

Creating a retinal implant, a visual cortical implant, or any other neural implant is a challenge for cognitive scientists, surgeons, electrical engineers, material scientists, biochemists, cell biologists, and computer scientists. For decades to come, scientists will be busy developing algorithms to emulate neural functions and control neuroprostheses;

semiconductor chips to implement the algorithms; microelectrode arrays to match the organic cytoarchitectonic scaffold of the nerves, spinal cord, or brain; and procedures to tune a device to a given patient. Ideally, human neural signals could be read remotely via transduction of electromechanical signals without surgically invading the body. These human-machine interfaces could be used to control a variety of external devices to improve human function and mobility.

Riener (chapter 22) reviews neurorehabilitation robotics and neuroprosthetics and shows how robots can be applied to support improved recovery in patients recovering from upper motor neuron and lower motor neuron lesions due to stroke, trauma, and other causes. Robots will make motion therapy more efficient by increasing training duration and number of training sessions, while reducing the number of therapists required per patient. Patient-cooperative control strategies have the potential to further increase the efficiency of robot-aided motion therapy. Neuroprosthetic function will be improved by applying closed-loop control components, computational models, and better BCIs.

Better BCIs will allow a user to interact with the environment without muscular activity (such as hand, foot, or mouth movement) and will require specific mental activity and strategy to modify brain signals in a predictive way. BCIs can be useful in augmented cognition applications (as described above) and may also help patients with paralysis due to amyotrophic lateral sclerosis, spinal cord injury, or other conditions. Pfurtscheller and colleagues (chapter 20) review BCIs that use EEG signals. The challenges are to record, analyze, and classify brain signals and transform them in real time into a control signal at the output of the BCI. This process is highly subject specific and requires rigorous training or learning sessions.

Guidelines, Standards, and Policy

Rapid development of devices, techniques, and applications in the field of neuroergonomics (neuroengineering, neural prostheses, augmented cognition, relational robots, and so on) have outpaced the development of guidelines, which are recommendations of good practice that rely on their authors' credibility and reputation, and of standards, which are formal documents published by standards-making bodies that are developed through extensive consultation, consensus, and formal voting. Guidelines and standards help establish best practices and provide a common frame of reference across new device designs over time to foster communication between different user groups, without hindering innovation. They allow reviewers to assess research quality, compare results across studies for quality and consistency, and identify inconsistencies and the need for further guidelines and standards.

Making fair guidelines and standards involves science, logic, policy, politics, and diplomacy to overcome entrenched interests and opinions, and to avoid favoritism toward or undue influence by certain persons or groups. It may be easier to propose standards than to apply them in practice, yet it is important to develop standards proactively before they become externally imposed. Relevant to neuroergonomics, the Food and Drug Administration is charged with administering standards for implantable brain devices, such as deep brain stimulators to mitigate Parkinson's disease, spinal cord stimulators for pain relief, cochlear implants for hearing, cardiac pacemakers to treat heart rhythm abnormalities, and vagal nerve stimulators to treat epilepsy or, lately, depression.

The Centers for Medicare and Medicaid Services will weigh in on efficacy and standards when it is asked to reimburse providers for rendering services that may as yet lack sufficient evidence to support clinical use (such as VR for phobias). Professional groups such as the American Academy of Neurology (AAN), American Academy of Neurosurgery, American Psychological Association, the Special Interest Group on Computer-Human Interaction of the Association for Computing Machinery, and the Human Factors and Ergonomics Society may also intervene to establish evidence-based quality standards when a practice, device, or treatment becomes professionally relevant and widespread.

For example, the Quality Standard Subcommittee of the AAN and similar groups in other medical subspecialties write clinical practice guidelines to assist their members in clinical decision making related to the prevention, diagnosis, prognosis, and treatment of medical disorders, which may come to include neuroergonomic applications or devices (such as neural implants to treat blindness or paralysis). The AAN guidelines make specific

practice recommendations based upon a rigorous review of all available scientific data. Key goals are to improve health outcomes, determine if practice follows current best evidence, identify research priorities based on gaps in the literature, promote efficient use of resources, and influence related public policy. Standards and guidelines in specific application areas discussed in this book, such as VR systems, have been published (Stanney, 2002). For a more comprehensive examination of standards across all areas of human factors and ergonomics, see Karwowski (2006).

Ethical Issues

Issues of privacy have been at the forefront since the advent of neuroergonomics (Hancock & Szalma, 2003) and are likely to continue to be so in the future too. Workers may be helped by automated systems that detect when fatigue, stress, or anxiety increase to levels that threaten performance and safety. But is there a dark side to such methods? For example, could those who seem especially stress or anxiety prone based on highly variable physiological measures and inaccurate predictive models be unfairly excluded from new opportunities or promotions? How will workers behave and what are their rights to privacy when it becomes possible to record seemingly everything they do, all the time, from multiple channels of data on brain and body states? The proliferation of embedded monitoring devices will expose events and behaviors that were once hidden behind cultural or institutional barriers. What could be done with these data beyond their intended purpose of enhancing safety, reducing stress, improving performance and health, and preventing injury?

In a similar vein, surveillance countermeasures to perceived terrorist threats include software intended to predict the intent of individuals bent on mayhem based on body movements, fidgeting, facial expression, eye glances, and other physiological indices. In chapter 13, Hancock and Szalma emphasize that we must be wary when private thoughts, opinions, and ideas—regardless of whether we like them or not—are unfairly threatened by other individuals, corporations, politicians, or arms of the state.

Remarkable ethical issues may unfold over the next century concerning the cooperative relationship between humans and machines at physical, mental, and social levels. How much should a person trust an automated adaptive assistant, avatar, virtual human, or affective computer that is in the loop with the human operator and potentially serving as a counselor, physician, friend, coworker, or supervisor? Who is the boss? How shall we mitigate concerns of control over human minds, bodies, and institutions by implants, robots, and software?

Modern discourse on somatic cell nuclear transfer (cloning) has sparked enormous hope and controversy at the borders between science, faith, policy, and politics. Emerging applications in neuroergonomics that are capable of reading human brain activity, predicting human failures, creating emotive robots, virtual human coworkers, companions, and bosses—and of hybridizing human and machine—may face similar trials (see also Clark, 2003). Nevertheless, we must face these challenges and not bury our heads in the sand like ostriches and hope they go away. Unless we (scientists and engineers) *ourselves* consider the ethical and privacy questions, others outside of science will decide the issues for us. The developments in neuroergonomics may lead to extraordinary opportunities for improved human-machine and human-human interaction. Others have noted that such developments may well represent the next major step in human evolution (Clark, 2003; Hancock & Szalma, 2003). Neuroergonomic technologies should be developed to serve humans and to help them engage in enjoyable and purposeful activity to ensure a well-evolved future.

Conclusion

As an interdisciplinary endeavor, neuroergonomics will continue to benefit from and grow alongside aggressive developments in neuroscience, ergonomics, psychology, engineering, and other fields. This ongoing synthesis will significantly advance our understanding of brain function underlying human performance of complex, real-world tasks. This knowledge can be put to use to design technologies for safer and more efficient operation in various work and home environments and in diverse populations of users. The basic enterprise of human factors and ergonomics—how humans design, interact with, and use technology—can be considerably enriched if we also consider the human brain that makes such activities possible. As the chapters in this

volume show, there have already been considerable achievements in basic research and application in neuroergonomics. The future is likely to yield more such advances.

References

Clark, A. (2003). *Natural born cyborgs: Minds, technologies, and the future of human intelligence*. New York: Oxford University Press.

De Pontbriand, R. (2005). Neuro-ergonomics support from bio- and nano-technologies. In *Proceedings of the Human Computer Interaction International Conference*. Las Vegas, NV: HCI International.

Gevins, A., Leong, H., Du, R., Smith, M., Le, J., DuRousseau, D., et al. (1995). Towards measurement of brain function in operational environments. *Biological Psychology, 40,* 169–186.

Hancock, P. A., & Szalma, J. L. (2003). The future of neuroergonomics. *Theoretical Issues in Ergonomics Science, 4*, 238–249.

Karwowski, W. (2006). *Handbook of standards and guidelines in ergonomics and human factors.* Mahwah, NJ: Erlbaum.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science, 1–2*, 5–20.

Stanney, K. (Ed.). (2002). *Handbook of virtual environments*. Mahwah, NJ: Erlbaum.

St. John, M., Kobus, D. A., Morrison, J. G., & Schmorrow, D. (2004). Overview of the DARPA augmented cognition technical integration experiment. *International Journal of Human-Computer Interaction, 17*, 131–149.

Vertegaal, R. (2002). Designing attentive user interfaces. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 23–30). New Orleans, Louisiana: SIGCHI.

10

Joel S. Warm and Raja Parasuraman

Cerebral Hemodynamics and Vigilance

The efficiency and safety of many complex human-machine systems can be critically dependent on the mental workload and vigilance of the operators of such systems. As pointed out by Wickens and Hollands (2000), it has long been recognized that the design of a high-quality human-machine system is not just a matter of assessing performance but also of evaluating how well operators can meet the workload demands imposed on them by the system. Major questions that must be addressed are whether human operators can meet additional unexpected demands when they are otherwise overloaded (Moray, 1979; Wickens, 2002) and whether they are able to maintain vigilance and respond effectively to critical events that occur at unpredictable intervals (Warm & Dember, 1998).

These considerations point to the need for sensitive and reliable measurement of human mental workload and vigilance. Behavioral measures, such as accuracy and speed of response to probe events, have been widely used to assess these psychological functions. However, as discussed by Kramer and Weber (2000), Parasuraman (2003), and Wickens (1990), measures of brain function offer some unique advantages that can be exploited in particular applications. Among these is the ability to extract covert physiological measures continuously in complex system operations in which overt behavioral measures may be relatively sparse. Perhaps a more compelling rationale is that measures of brain function can be linked to emerging cognitive neuroscience knowledge on attention (Parasuraman & Caggiano, 2002; Posner, 2004), thereby allowing for the development of neuroergonomic theories that in turn can advance practical applications of research on mental workload and vigilance.

In this chapter, we describe a series of recent neuroergonomic studies from our research group on vigilance, focusing on the use of noninvasive measurement of cerebral blood flow velocity. We use a theoretical framework of attentional resources (Kahneman, 1973; Moray, 1967; Navon & Gopher, 1979; Norman & Bobrow, 1975; Posner & Tudela, 1997; Wickens, 1984). Resource theory is the dominant theoretical approach to the assessment of human mental workload (Wickens, 2002) and also provides a major conceptual framework for understanding human vigilance performance (Parasuraman, 1979; Warm & Dember, 1998). Consistent with the view first proposed by Sir Charles Sherrington (Roy & Sherrington, 1890), a considerable amount of research on brain imaging indicates that there is a close tie between cerebral blood flow and neural activity in the performance of mental tasks

(Raichle, 1998; Risberg, 1986). Consequently, changes in blood flow velocity and oxygenation in our studies are considered to reflect the availability and utilization of information processing assets needed to cope with the vigilance task.

The Hemodynamics of Vigilance

Brain Systems and Vigilance

Vigilance involves the ability of observers to detect transient and infrequent signals over prolonged periods of time. That aspect of human performance is of considerable concern to human factors and ergonomic specialists because of the critical role that vigilance occupies in many complex human-machine systems, including military surveillance, air-traffic control, cockpit monitoring and airport baggage inspection, industrial process and quality control, and medical functions such as cytological screening and vital sign monitoring during surgery (Hancock & Hart, 2002; Parasuraman, 1986; Warm & Dember, 1998). Thus, it is important to understand the neurophysiological factors that control vigilance performance.

In recent years, brain imaging studies using positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) techniques have been successful in demonstrating that changes in cerebral blood flow and glucose metabolism are involved in the performance of vigilance tasks (see review by Parasuraman, Warm, & See, 1998). These studies have also identified several brain regions that are active in such tasks, including the right frontal cortex and the cingulate gyrus, as well as subcortical nuclei such as the locus coeruleus. Although these studies have identified brain regions involved in vigilance, Parasuraman et al. (1998) have pointed out some major limitations of this research. With the exception of PET studies by Paus et al. (1997) and by Coull, Frackowiak, and Frith (1998), the brain imaging studies have neglected to link the systems they have identified to performance efficiency, perhaps because of the high cost associated with using PET and fMRI during the prolonged running times characteristic of vigilance research. Thus, the functional role of the brain systems identified in the imaging studies remains largely unknown. Gazzaniga, Ivry, and Mangun (2002) have also emphasized the necessity of linking neuroimaging results to human performance for enhanced understanding of research in cognitive neuroscience.

Other problems with the PET and fMRI procedures are that they feature restrictive environments in which observers need to remain almost motionless throughout the scanning procedure so as not to compromise the quality of the brain images, and fMRI acquisition is accompanied by loud noise. Observers in vigilance experiments rarely remain motionless, however. Instead, research has shown that they tend to fidget during the performance of a vigilance task, with the amount of motor activity increasing with time on task (Galinsky, Rosa, Warm, & Dember, 1993). Moreover, the noise of fMRI is one of several environmental variables that can degrade vigilance performance. For example, Becker, Warm, Dember, and Hancock (1995) showed that noise lowered perceptual sensitivity in a vigilance task, interfered with the ability of observers to profit from knowledge of results, and elevated perceived mental workload. Accordingly, the conditions required for the effective use of the PET and fMRI techniques may not provide a suitable environment for linking changes in brain physiology with vigilance performance over a prolonged period of time. To meet this need, we turned to two other imaging procedures, transcranial Doppler sonography (TCD) and transcranial cerebral oximetry.

Transcranial Doppler Sonography

TCD is a noninvasive neuroimaging technique that employs ultrasound signals to monitor cerebral blood flow velocity or hemovelocity in the mainstem intracranial arteries—the middle, anterior, and posterior arteries. These arteries are readily isolated through a cranial "transtemporal window" and exhibit discernible measurement characteristics that facilitate their identification (Aaslid, 1986). The TCD technique uses a small 2 MHz pulsed Doppler transducer to gauge arterial blood flow. The transducer is placed just above the zygomatic arch along the temporal bone, a part of the skull that is functionally transparent to ultrasound. The depth of the pulse is adjusted until the desired intracranial artery (e.g., the middle cerebral artery, MCA) is isonated. TCD measures the difference in frequency between the outgoing and reflected energy as it strikes moving erythrocytes.

The low weight and small size of the transducer and the ability to embed it conveniently in a headband permit real-time measurement of cerebral blood flow while not limiting, or being hampered by, body motion. Therefore, TCD enables inexpensive, continuous, and prolonged monitoring of cerebral blood flow velocity concurrent with task performance. Blood flow velocities, measured in centimeters per second, are typically highest in the MCA, and the MCA carries about 80% of the blood flow within each cerebral hemisphere (Toole, 1984). Consequently, our TCD studies of mental workload and vigilance assess blood flow velocity in the MCA, but other TCD studies, particularly those examining perceptual processes, also measure blood flow in the posterior cerebral artery (PCA). For further methodological details of the TCD technique, see chapter 6.

When a particular area of the brain becomes metabolically active, as in the performance of mental tasks, by-products of this activity such as carbon dioxide (CO_2) increase. This increase in CO_2 leads to a dilation of blood vessels serving that area, which in turn results in blood flow to that region (Aaslid, 1986). Consequently, TCD offers the possibility of measuring changes in metabolic activity during task performance. The use of TCD in brain imaging performance applications is limited, in part, by its low spatial resolution: TCD can supply gross hemispheric data, but it does not provide information about changes in specific brain loci, as is the case with PET and fMRI. Nevertheless, TCD offers good temporal resolution (Aaslid, 1986) and, compared to PET and fMRI, it can track rapid changes in blood flow dynamics that can be followed in real time under less restrictive and invasive conditions. The use of TCD to index blood flow changes in a wide variety of cognitive, perceptual, and motor tasks has been reviewed elsewhere (Duschek & Schandry, 2003; Klingelhofer, Sander, & Wittich, 1999; Stroobant & Vingerhoets, 2000; see also, chapter 6).

Transcranial Cerebral Oximetry

The TCD technique provides a very economical way to assess cerebral blood flow in relatively unrestricted environments. However, TCD does not directly provide information on oxygen utilization in the brain, which would be useful to assess as another indicator of the activation of neuronal populations recruited in the service of cognitive processes. Optical imaging, in particular near-infrared spectroscopy (NIRS), can be used in the assessment of cerebral oximetry. There are several types of NIRS technology, including the recent development of so-called fast NIRS, as discussed in chapter 5. The standard NIRS technique has several advantages over TCD, including the ability to assess activation in several brain regions, and not just in the left and right hemispheres as with TCD. Previous research using NIRS has shown that tissue oxygenation increases with the information processing demands of the task being performed by an observer (Punwani, Ordidge, Cooper, Amess, & Clemence, 1998; Toronov et al., 2001). Hence, one might expect that along with cerebral blood flow, cerebral oxygenation would also be related to both mental workload and to vigilance.

TCD Studies of Vigilance

Working Memory and Vigilance

The initial study from our research group to examine TCD in relation to vigilance was conducted by Mayleben (1998). That study was guided by the finding that working memory demand can be a potent influence on vigilance performance (Davies & Parasuraman, 1982; Parasuraman, 1979). Parasuraman (1979) had first showed that successive discrimination tasks, in which the detection of critical targets requires comparison of information in working memory, are more susceptible to performance decrement over time than simultaneous discrimination tasks, which have no such memory imperative. The role of memory representation in the vigilance decrement was confirmed in a study by Caggiano and Parasuraman (2004). Moreover, Warm and Dember (1998) conducted a series of studies showing that other psychophysical and task factors that reduce attentional resources (e.g., low signal salience, dual-task demands) have a greater detrimental effect on successive than on simultaneous vigilance tasks. These findings can be interpreted in terms of the resource model described earlier in this chapter. According to that model, a limited-capacity information processing system allocates resources to cope with situations that confront it. The vigilance

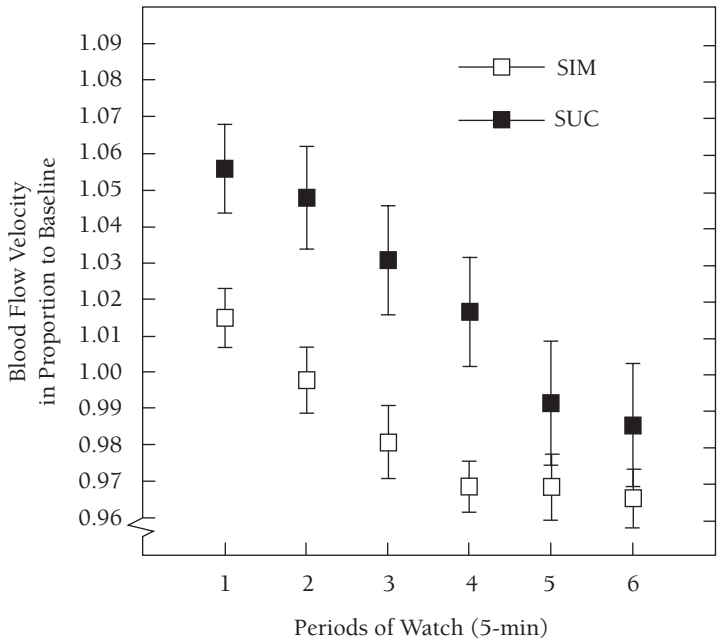


Figure 10.1. Mean cerebral blood flow velocity as a function of periods of watch for simultaneous-type (SIM) and successive-type (SUC) vigilance tasks. Error bars are standard errors. After Mayleben (1998).

decrement, the decline in signal detections over time that characterizes vigilance performance (Davies & Parasuraman, 1982; Warm & Jerison, 1984), reflects the depletion of information processing resources or reservoirs of energy that cannot be replenished in the time available. Given that changes in blood flow might reflect the availability and utilization of the information processing assets needed to cope with a vigilance task, Mayleben (1998) hypothesized that the vigilance decrement should be accompanied by a decline in cerebral hemovelocity and that the overall level of blood flow should be greater for a memory-demanding successive task than for a memory-free simultaneous task.

Participants in this study were asked to perform either a successive or a simultaneous vigilance task during a 30-minute vigil. Critical signals for detection in the simultaneous task were cases in which one of two lines on a visual display was slightly taller than the other. In the successive task, critical signals were cases in which both lines were slightly taller than usual. Pilot work ensured that the tasks were equated for difficulty under alerted conditions. In this and in all of the subsequent studies from our laboratory described in this chapter, blood flow or hemovelicity is expressed as a percentage of the last 60 seconds of a 5-minute resting baseline, as recommended by Aaslid (1986). As illustrated in figure 10.1, Mayleben (1998) found that the vigilance decrement in detection rate over time was accompanied by a parallel decline in cerebral hemovelocity. Also consistent with expectations from a resource model, the overall level of blood flow velocity was significantly higher for observers who performed the successive task than for those who performed the simultaneous task.

An important additional finding of the Mayleben (1998) study was that the blood flow effects were lateralized—hemovelocity was greater in the right than in the left hemisphere, principally in the performance of the memory-based successive task. A result of this sort is consistent with earlier PET and psychophysical studies showing right-brain superiority in vigilance (Parasuraman et al., 1998) and with studies by Tulving, Kapur, Craik, Moscovitch, and Houle (1994) indicating that memory retrieval is primarily a right-brain function. Schnittger, Johannes, Arnavaz, and Munte (1997) also reported the performance–blood flow relation over time described in the Mayleben (1998) study.

However, a clear coupling of blood flow and performance could not be determined in their investigation because of the absence of a control for the possibility of spontaneous declines in blood flow over time, such as may result from systemic declines in arousal. Following a suggestion by Parasuraman (1984), Mayleben (1998) employed such a control by exposing a group of observers to the dual-line display for 30 minutes in the absence of a work imperative. Blood flow remained stable over the testing period under such conditions. Thus, the decline in cerebral blood flow was closely linked to the need to maintain attention to the visual display and not merely to the passage of time.

A potential challenge to an interpretation of these results along resource theory lines comes from the findings that blood flow velocity is sensitive to changes in blood pressure and cardiac output (Caplan et al., 1990) and that changes in heart rate variability are correlated with vigilance performance (Parasuraman, 1984). Accordingly, one could argue that the performance and hemovelocity findings in this study do not reflect information processing per se but rather a gross change in systemic vascular activity that covaried with blood flow. The lateralization of the performance and hemovelocity findings challenges such a view, since gross changes in vascular activity are not likely to be hemisphere dependent.

Controlling the Vigilance Decrement with Signal Cueing

Signal detection in vigilance can be improved by providing observers with consistent and reliable cues to the imminent arrival of critical signals. As previous experiments have shown, the principal consequence of such forewarning is the elimination of the vigilance decrement (Annett, 1996; Wiener & Attwood, 1968). The cueing effect can be linked to resource theory as follows. Observers need to monitor a display only after having been prompted about the arrival of a signal, and therefore can husband their information processing resources over time. In contrast, when no cues are provided, observers are never certain of when a critical signal might appear, and consequently have to process information on their displays continuously across the watch, thereby consuming more of their resources over time than cued observers. Thus, one can predict that in the presence of perfectly reliable cueing, the temporal decline in cerebral blood flow would be attenuated in comparison to a noncued condition and also in comparison to conditions in which cueing was less than perfectly reliable, since observers in such conditions would not be relieved of the need to attend continuously to the vigilance display.

This prediction was tested by Hitchcock et al. (2003) using a simulated air-traffic control (ATC) display. Critical signals for detection were pairs of aircraft traveling on a potential collision course. Observers monitored the simulated ATC display for 40 minutes. To manipulate perceptual difficulty, signal salience, varied by changing the Michaelson contrast ratio of the aircraft to their background, was high (98%, dark black aircraft on a light background) or low (2%, light gray aircraft on a light background). Signal salience was combined factorially with four levels of cue reliability—100% reliable, 80% reliable, 40% reliable, and a no-cue control. Observers in the cueing groups were instructed that a critical signal would occur within one of the five display updates immediately following the verbal prompt *look* provided through a digitized male voice. Observers in each of the cue groups were advised about the reliability of the cues they would receive. To control for accessory auditory stimulation, observers in the no-cue group received acknowledgment after each response in the form of the word *logged* spoken in the same male voice.

As can be seen in figure 10.2, the detection scores for the several cueing conditions in this study were similar to each other during the early portion of the vigil and diverged by the end of the session. More specifically, performance efficiency remained stable in the 100% reliable cueing condition but declined over time in the remaining conditions, so that by the end of the vigil, performance efficiency was clearly best in the 100% group followed in order by the 80%, 40%, and no-cue groups.

The hemovelocity scores from the left hemisphere showed a significant decline over time but no effect for cueing with either high-salience or low-salience signals. A similar result was found for high-salience signals in the right hemisphere. Cueing effects emerged, however, in the hemovelocity scores for the right hemisphere with low-salience signals. As was the case with detection probability, the hemovelocity scores for the several cueing conditions were similar to each other during the early portions of the vigil, but showed differential rates

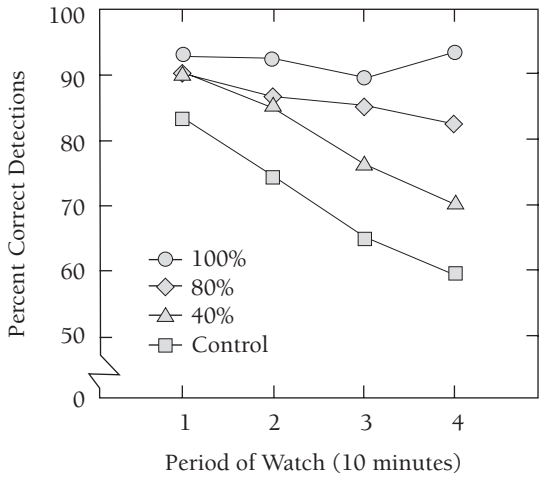


Figure 10.2. Percentages of correct detections as a function of periods of watch for four cue-reliability conditions. After Hitchcock et al. (2003).

of decline over time, so that by the end of the vigil, blood flow was clearly highest in the 100% group followed in order by the 80%, 40%, and no cue groups. This result is illustrated in figure 10.3.

In summary, the hemovelocity scores taken from the right hemisphere under low salience almost exactly mirrored the effects of cuing on performance efficiency. The finding that this result was limited to low-salience signals is consistent with a study by Korol and Gold (1998) indicating that brain systems involving glucose metabolism need to be sufficiently challenged in order for measurable physiological changes to emerge in cognitive and attentional processing tasks. Restriction of the cue-time-salience hemovelocity findings to the right hemisphere is consistent with expectations about right hemisphere control of vigilance. As in the initial study, blood flow remained stable over time in both hemispheres throughout the watch when observers were exposed to the simulated air-traffic display without a work imperative.

Visual Search

The ATC task used in the Hitchcock et al. (2003) study was such that critical signals could be detected without any substantial need for searching the display. In contrast, many real-world environments, both in ATC and elsewhere, require that operators conduct a visual search of displays in order to detect critical signals. A well-established finding from laboratory studies of visual search is the *search asymmetry* phenomenon (Treisman & Gormican, 1988). This effect refers to more rapid detections

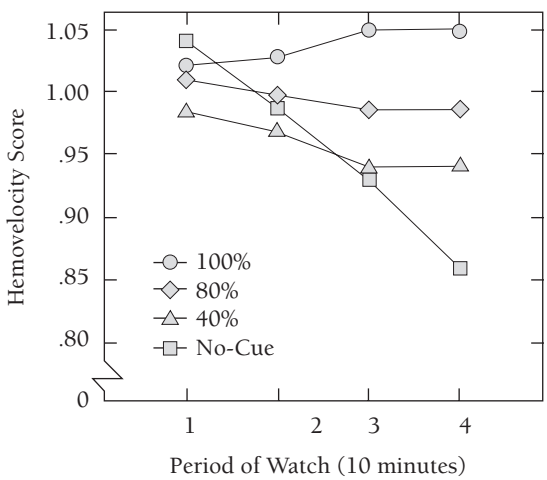


Figure 10.3. Mean hemovelocity scores as a function of periods of watch for four cue-reliability conditions. Data are from the right hemisphere/low-signal salience condition. After Hitchcock et al. (2003).

when searching for the presence of a distinguishing feature in an array of stimuli as opposed to its absence. Indeed, when searching for presence, the distinguishing feature appears to be so salient that it seems to pop out of the display. The phenomenon of search asymmetry has been accounted for by the feature integration model (Treisman & Gormican), which suggests that searching for the presence of a feature is guided by preattentive, parallel processing, while more deliberate serial processing is required for determining its absence.

Studies by Schoenfeld and Scerbo (1997, 1999) have extended the presence-absence distinction to the accuracy of signal detections in long-duration sustained attention or vigilance tasks. Performance efficiency in vigilance tasks varies inversely with the information processing demand imposed by the task, as indexed by the number of stimulus elements that must be scanned in search of critical signals (Grubb, Warm, Dember, & Berch, 1995; Parasuraman, 1986). The view that detecting the absence of a feature is more capacity demanding than detecting its presence led Schoenfeld and Scerbo (1997, 1999) to predict that increments in the number of array elements to be scanned in separating signals from noise in vigilance would have a more negative effect upon signal detection in the feature absence than in the feature presence case. Consistent with that prediction, they found that when observers were required to detect the absence of a figure, signal detectability declined as the size of the stimulus array to be scanned was increased from two to five elements. Increasing array size, however, had no effect on performance when observers were required to monitor for the presence of that feature. In addition, observers rated the workload of their assignment to be greater when monitoring for feature absence than presence on the NASA Task Load Index (TLX) scale, a standard subjective report measure of the perceived mental workload imposed by a task (Hart & Staveland, 1988; Warm, Dember & Hancock, 1996; Wickens & Hollands, 2000).

In most vigilance tasks, critical signals for detection are embedded in a background of repetitive nonsignal or neutral events. Several studies have demonstrated that signal detections vary inversely with the background event rate and that this effect is more prominent in tasks requiring high information processing demand than low (Lanzetta, Dember, Warm, & Berch, 1987; Parasuraman, 1979; Warm & Jerison, 1984). Given that detecting the absence of a feature is more capacity demanding than detecting its presence, Hollander and his associates (2004) hypothesized that the degrading effects of increments in background event rate would be more pronounced when observers monitored for the absence than for the presence of a feature. As in the studies by Schoenfeld and Scerbo (1997, 1999), perceived mental workload was also anticipated to be greater when observers monitored for feature absence than for feature presence. With regard to blood flow, Hollander et al. (2004) predicted that blood flow would be higher when observers were required to detect feature absence than presence and would show a greater decline over time in the absence than in the presence condition. The two types of tasks, presence and absence, were combined factorially with three levels of event rate, 6, 12, and 24 events per minute, to provide six experimental conditions. In all conditions, observers participated in a 40-minute vigil divided into four 10-minute periods of watch during which they monitored an array of five circles positioned around the center of a video display terminal at the 3, 5, 7, 9, and 12 o'clock locations. The critical signal for detection in the presence condition was the appearance of a vertical 4 mm line intersecting the 6 o'clock position within one of the circles in the array. In the absence condition, the vertical line was present in all circles but one. Ten critical signals were presented in each watchkeeping period in all experimental conditions.

As anticipated, the event rate effect was indeed more pronounced in the absence than in the presence condition. Detection probability remained stable across event rates in the presence condition but declined with increments in event rate in the absence condition. Another important aspect of the performance data was the finding that signal detections declined significantly over time in both the feature presence and absence conditions.

The finding that the event rate effect was more pronounced when observers had to monitor for stimulus absence than presence is reminiscent of the findings in the earlier reports by Schoenfeld and Scerbo (1997, 1999) that the effect of another information processing factor in vigilance, the size of the element set that must be scanned in search of critical signals, is also more notable in the feature absence than the feature presence condition. Also consistent with Schoenfeld and Scerbo (1997, 1999), perceived mental workload was greater

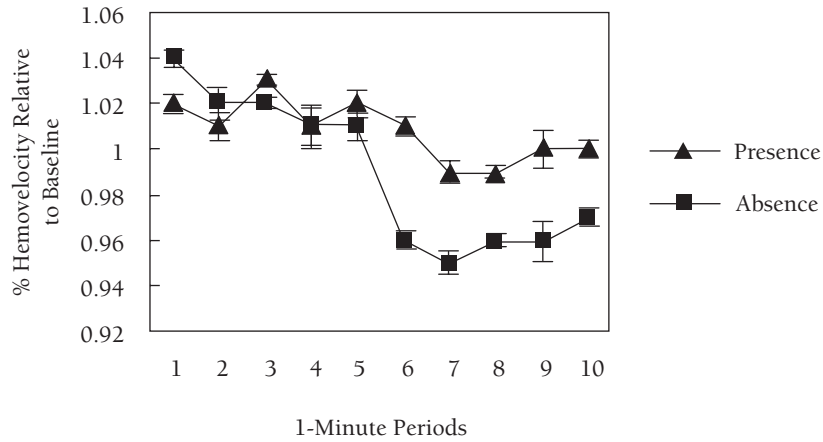


Figure 10.4. Mean hemovelocity scores in feature presence or absence conditions for successive 1-minute intervals during the initial period of watch. Data are for the right hemisphere. Error bars are standard errors. After Hollander et al. (2004).

when participants monitored for feature absence than presence. Thus, the results support the notion of differential capacity demand in detecting feature absence than presence. However, the finding of a vigilance decrement in the feature presence condition suggests that counter to the early claim in feature-integration theory that feature detection is preattentive, some information processing cost must be associated with detecting feature presence. This interpretation is supported by the fact that even though the mental workload of the presence condition was less than that of the absence condition, it still fell in the upper level of the NASA TLX scale. Spatial cueing studies in which the size of a precue is varied prior to the presentation of the search display have also found cue-size effects on target identification time for both feature and conjunction search (Greenwood & Parasuraman, 1999, 2004). Results such as these are consistent with the emerging view in the search literature that the alignment of feature detection with purely preattentive processing may no longer be tenable (Pashler, 1998; Quinlan, 2003).

When measured over the 10-minute intervals of watch, blood flow in the Hollander et al. (2004) study was found to be greater in the presence than in the absence condition. That result was counter to expectations based on the view that detecting feature absence is more capacity demanding than detecting feature presence. It is conceivable, however, that this apparent reversal of the expected effect reflected the fact than the demands of feature absence were great enough to tax information processing resources very early in the vigil and that those resources were not replenished over time. An account along that line would be supported if it could be shown that while differences in blood flow are greater in the feature absence than in the feature presence case at the very outset of the vigil, the reverse effect emerged as the vigil continued. Toward that end, a fine-grained minute-by-minute analysis was performed on the blood flow scores of the presence and absence conditions during the initial watchkeeping period in the left and right hemispheres. No task differences were noted for the left hemisphere. However, as shown in figure 10.4, blood flow velocity in the right hemisphere was greater in the feature absence than in the feature presence condition at the outset of the vigil, and the reverse effect emerged after observers had performed the task for 6 minutes. Statistical tests indicated that there were no significant differences between the two tasks in the first 5 minutes of watch but that there were statistically significant differences between the conditions in the 6th through the 10th minutes of watch. Similar fine-grained examination of the data for the remaining periods of watch revealed that the reduced level of blood flow in the absence condition that emerged halfway through the initial watchkeeping period remained consistent throughout each minute of all of the remaining periods of watch.

Evidently, Hollander et al.'s (2004) initial expectation of greater blood flow in the absence condition

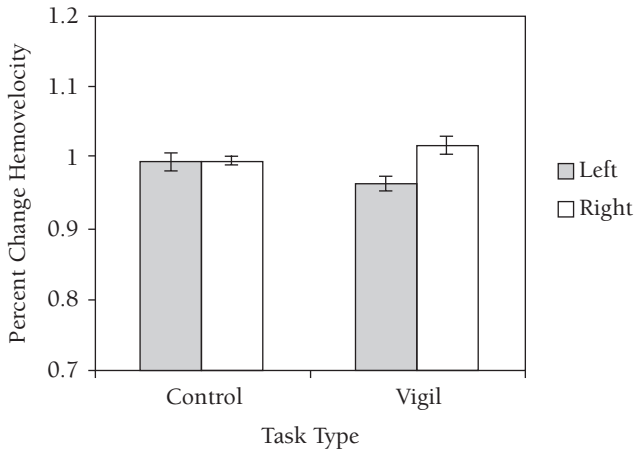


Figure 10.5. Percentage change in hemovelocity relative to resting baseline in the left and right cerebral hemispheres for the control and active vigilance conditions. After Helton et al. (in press).

underestimated the degree to which that condition taxed information processing resources in the vigilance task. Rather than being reflected in an overall elevation in blood flow, the greater information processing demand exerted by the absence condition was evident in an early-appearing drain on resources. Although initially unanticipated, this finding from the fine-grained analysis of the data is consistent with the expectation of a steeper decline in blood flow associated with the absence condition. As in the Hitchcock et al. (2003) and Mayleben (1998) studies, hemovelocity in the Hollander et al. (2004) study remained stable over the course of the watch in both hemispheres among control observers who viewed the displays without a work imperative, indicating once again that the blood flow effects were indeed task dependent.

The Abbreviated Vigil

Thus far we have discussed TCD findings in studies that made use of traditional vigilance tasks lasting 30 minutes or more. Because of their long duration, investigators have found it inconvenient to incorporate such tasks in test batteries or, as discussed previously, to link vigilance performance with brain imaging metrics such as PET and fMRI. Accordingly, it is of interest to examine whether the TCD-vigilance findings can be replicated in shorter-duration vigilance tasks (Nuechterlein, Parasuraman, & Jiang, 1983; Posner, 1978; Temple et al., 2000). Toward that end, Helton et al. (in press) used a 12-minute vigilance task developed by Temple et al. (2000) in which participants were asked to inspect the repetitive presentation on a VDT of light gray capital letters consisting of an O, a D, and a backward D. The letters were presented for only 40 ms at a rate of 57.5 events per minute and exposed against a visual mask consisting of unfilled circles on a white background. Critical signals for detection were the appearances of the letter O (signal probability = 0.20/period of watch). In addition to TCD measurement of blood flow, Helton et al. (in press) also employed the NIRS procedure to measure transcranial cerebral oximetry via a Somanetics INVOS 4100 Cerebral Oximeter.

As can be seen in figures 10.5 and 10.6, both blood flow velocity (figure 10.5) and oxygenation (figure 10.6) were found to be significantly higher in the right than in the left cerebral hemisphere among observers who performed the vigilance task while there were no hemispheric differences in the blood flow and oxygenation measures among control observers who viewed the vigilance display without a work imperative. In this study, the oxygenation measure was based on a percentage of a 3-minute resting baseline.

Clearly, the results of this study indicated that performance in the abbreviated vigil was right lateralized, a finding that coincides with the outcome of earlier blood flow studies featuring more traditional long-duration vigils and with PET and fMRI investigations (Parasuraman et al., 1998). This parallel has several important implications. It provides strong support for the argument that the abbreviated vigil is a valid analog of long-duration vigilance tasks. The fact that the NIRS procedure yielded laterality effects similar to those of the TDC

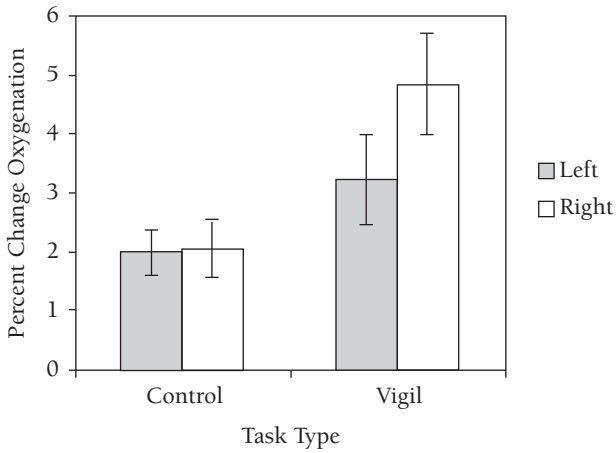


Figure 10.6. Percentage change frontal lobe oxygenation relative to resting baseline in the left and right cerebral hemispheres for the control and active vigilance conditions. After Helton et al. (in press).

procedure further implies that laterality in vigilance is a generalized effect that appears in terms of both hemovelocity and blood oxygenation. It also implies that the NIRS procedure may be a useful supplement to the TCD approach in providing a noninvasive imaging measure of brain activity in the performance of a vigilance task.

It is important to note that while Helton et al.'s (in press) results regarding laterality of function with the abbreviated vigil were consistent with those found with its long-duration analogs, their findings regarding the decrement function were not. Both the TCD and the NIRS indices remained stable over the course of the watch while performance efficiency declined over time. It is possible that cerebral hemodynamics are structured so that overall hemispheric dominance emerges early in the time course of task performance but that temporally based declines in cerebral blood flow and blood oxygen levels require a considerable amount of time to become observable. Thus, the abbreviated 12-minute vigil, which is only about 30% as long as the vigils employed in earlier vigilance studies, may not be long enough to permit time-based declines in blood flow or blood oxygen levels.

Conclusion

One of the goals of neuroergonomics is to enhance understanding of aspects of human performance in complex systems with respect to the underlying brain mechanisms and to provide measurement tools to study these mechanisms (Parasuraman, 2003). From this perspective, the use of TCD-based measures of cerebral blood flow to assess human mental workload and vigilance can be considered a success. The vigilance studies have revealed a close coupling between vigilance performance and blood flow and they provide empirical support for the notion that blood flow may represent a metabolic index of information processing resource utilization during sustained attention. The demonstration of systematic modulation of blood flow in the right cerebral hemisphere with time on task, memory load, signal salience and cueing, the detection of feature absence or presence, and target detection in Temple et al.'s (2000) abbreviated vigil provides evidence for a right hemispheric brain system that is involved in the functional control of vigilance performance over time.

Another goal of neuroergonomics research is to use knowledge of brain function to enhance human-system performance. In additional to the theoretical and empirical contributions of TCD research, there are also some potentially important ergonomic ramifications. TCD may offer a noninvasive and inexpensive tool to "monitor the monitor" and to help decide when operator vigilance has reached a point where task aiding is necessary or operators need to be rested or removed. NIRS-based measurement of blood oxygenation may provide similar information.

Main Points

1. Transcranial Doppler sonography and near-infrared spectroscopy can be used to measure cerebral blood flow velocity and cerebral

oxygenation, respectively, during the performance of a vigilance task.
2. The temporal decline in signal detections that characterizes vigilance performance is accompanied by a similar decline in brain blood flow over time.
3. Brain blood flow is greater in the performance of a memory-demanding successive-type vigilance task than for a memory-free simultaneous-type task.
4. Changes in signal detection in vigilance brought about by variations in cue reliability are paralleled by changes in brain blood flow.
5. The temporal decline in blood flow accompanying vigilance performance is greater when critical signals for detection are defined by the absence than by the presence of a target element.
6. Both the TCD and NIRS measures point to a right hemispheric system in the control of vigilance performance.

Key Readings

Aaslid, R. (1986). Transcranial Doppler examination techniques. In R. Aaslid (Ed.), *Transcranial Doppler sonography* (pp. 39–59). New York: Springer-Verlag.

Duschek, S., & Schandry, R. (2003). Functional transcranial Doppler sonography as a tool in psychophysiological research. *Psychophysiology, 40,* 436–454.

Hitchcock, E. M., Warm, J. S., Matthews, G., Dember, W. N., Shear, P. K., Tripp, L. D., et al. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science*, *4*, 89–112.

Parasuraman, R., Warm, J. S., & See, J. W. (1998). Brain systems of vigilance. In R. Parasuraman (Ed.), *The attentive brain* (pp. 221–256). Cambridge, MA: MIT Press.

References

Aaslid, R. (1986). Transcranial Doppler examination techniques. In R. Aaslid (Ed.), *Transcranial Doppler sonography* (pp. 39–59). New York: Springer-Verlag.

Annett, J. (1996). Training for perceptual skills. *Ergonomics*, *9*, 459–468.

Becker, A. B., Warm, J. S., Dember, W. N., & Hancock, P. A. (1995). Effects of jet engine noise and performance feedback on perceived workload in a monitoring task. *International Journal of Aviation Psychology, 5*, 49–62.

Caggiano, D. M., & Parasuraman, R. (2004). The role of memory representation in the vigilance decrement. *Psychonomic Bulletin and Review, 11*, 932–937.

Caplan, L. R., Brass, L. M., DeWitt, L. D., Adams, R. J., Gomex, C., Otis, S., et al. (1990). Transcranial Doppler ultrasound: Present status. *Neurology*, *40*, 496–700.

Coull, J. T., Frackowiak, R. J., & Frith, C. D. (1998). Monitoring for target objects: Activation of right frontal and parietal cortices with increasing time on task. *Neuropsychologia*, *36*, 1325–1334.

Davies, D. R., & & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Duschek, S., & Schandry, R. (2003). Functional transcranial Doppler sonography as a tool in psychophysiological research. *Psychophysiology, 40*, 436–454.

Galinsky, T. L., Rosa, R. R., Warm, J. S., & Dember, W. N. (1993). Psychophysical determinants of stress in sustained attention. *Human Factors*, *35*, 603–614.

Gazzaniga, M. S., Ivry, R., & Mangun, G. R. (2002). *Cognitive neuroscience: The biology of the mind* (2nd ed.). New York: Norton.

Greenwood, P. M., & Parasuraman, R. (1999). Scale of attentional focus in visual search. *Perception & Psychophysics, 61*, 837–859.

Greenwood, P. M., & Parasuraman, R. (2004). The scaling of spatial attention in visual search and its modification in healthy aging. *Perception & Psychophysics*, *66*, 3–22.

Grubb, P. L., Warm, J. S., Dember, W. N. & Berch, D. B. (1995). Effects of multiple signal discrimination on vigilance performance and perceived workload. *Proceedings of the Human Factors and Ergonomics Society, 39th annual meeting*, 1360–1364.

Hancock, P. A., & Hart, G. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics and Design*, *10*, 6–16.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North Holland.

Helton, W. S., Hollander, T. D., Warm, J. S., Tripp, L. D., Parsons, K., Matthews, G., et al. (in press). The abbreviated vigilance task and cerebral hemody-

namics. *Journal of Clinical and Experimental Neuropsychology.*

Hitchcock, E. M., Warm, J. S., Matthews, G., Dember, W. N., Shear, P. K., Tripp, L. D., et al. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science*, *4*, 89–112.

Hollander, T. D., Warm, J. S., Matthews, G., Shockley, K., Dember, W. N., Weiler, E. M., et al. (2004). Feature presence/absence modifies the event rate effect and cerebral hemovelocity in vigilance. *Proceedings of the Human Factors and Ergonomics Society, 48th annual meeting*, 1943–1947.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Klingelhofer, J., Sander, D., & Wittich, I. (1999). Functional ultrasonographicic imaging. In V. L. Babikian & L. R. Wechsler (Eds.), *Transcranial Doppler ultrasonography* (2nd ed., pp. 49–66). Boston: Butterworth Heinemann.

Korol, D. L., & Gold, P. E. (1998). Glucose, memory, and aging. *American Journal of Clinical Nutrition*, *67*, 764–771.

Kramer, A. F., & Weber, T. (2000). Applications of psychophysiology to human factors. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 794–814). New York: Cambridge University Press.

Lanzetta, T. M., Dember, W. N., Warm, J. S., & Berch, D. B. (1987). Effects of task type and stimulus homogeneity on the event rate function in sustained attention. *Human Factors, 29*, 625–633.

Mayleben, D. W. (1998). *Cerebral blood flow velocity during sustained attention.* Unpublished doctoral dissertation, University of Cincinnati, OH.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica, 27*, 84–92.

Moray, N. (1979). *Mental workload.* New York: Plenum.

Navon, D., & Gopher, D. (1979). On the economy of human processing systems. *Psychological Review*, *86*, 214–255.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44–64.

Nuechterlein, K., Parasuraman, R., & Jiang, Q. (1983). Visual sustained attention: Image degradation produces rapid sensitivity decrement over time. *Science, 220*, 327–329.

Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science, 205*, 924–927.

Parasuraman, R. (1984). The psychobiology of sustained attention. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 61–101). London: Wiley.

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception: Vol. 2. Cognitive processes and performance* (pp. 43.1–43.39). New York: Wiley.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science, 4*, 5–20.

Parasuraman, R., & Caggiano, D. (2002). Mental workload. In V. S. Ramachandran (Ed.), *Encyclopedia of the human brain* (Vol. 3, pp. 17–27). San Diego: Academic Press.

Parasuraman, R., Warm, J. S., & See, J. W. (1998). Brain systems of vigilance. In R. Parasuraman (Ed.), *The attentive brain* (pp. 221–256). Cambridge, MA: MIT Press.

Pashler, H. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Paus, T., Zatorre, R. J., Hofle, N., Caramanos, Z., Gotman, J., Petrides, M., et al. (1997). Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task. *Journal of Cognitive Neuroscience, 9*, 392–408.

Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

Posner, M. I. (2004). *Cognitive neuroscience of attention.* New York: Guilford.

Posner, M. I., & Tudela, P. (1997). Imaging resources. *Biological Psychology, 45*, 95–107.

Punwani, S., Ordidge, R. J., Cooper, C. E., Amess, P., & Clemence, M. (1998). MRI measurements of cerebral deoxyhaemoglobin concentration (dhB)—correlation with near infrared spectroscopy (NIRS). *NMR in Biomedicine*, *11*, 281–289.

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin, 129*, 643–673.

Raichle, M. E. (1998). Behind the scenes of functional brain imaging: A historical and physiological perspective. *Proceedings of the National Academy of Sciences, USA, 95*, 765–772.

Risberg, J. (1986). Regional cerebral blood flow in neuropsychology. *Neuropsychologica, 34*, 135–140.

Roy, C. S., & Sherrington, C. S. (1890). On the regulation of the blood supply of the brain. *Journal of Physiology (London), 11,* 85–108.

Schnittger, C., Johannes, S., Arnavaz, A., & Munte, T. F. (1997). Relation of cerebral blood flow velocity and level of vigilance in humans. *NeuroReport*, *8*, 1637–1639.

Schoenfeld, V. S., & Scerbo, M. W. (1997). Search differences for the presence and absence of features in sustained attention. *Proceedings of the Human*

Factors and Ergonomics Society, 41st annual meeting, 1288–1292.

Schoenfeld, V. S., & Scerbo, M. W. (1999). The effects of search differences for the presence and absence of features on vigilance performance and mental workload. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 177–182). Mahwah, NJ: Erlbaum.

Stroobant, N., & Vingerhoets, G. (2000). Transcranial Doppler ultrasonography monitoring of cerebral hemodynamics during performance of cognitive tasks: A review. *Neuropsychology Review, 10*, 213–231.

Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors*, *42*, 183–194.

Toole, J. F. (1984). *Cerebralvascular disorders* (3rd ed.). New York: Raven.

Toronov, V., Webb, A., Choi, J. H., Wolf, M., Michalos, A., Gratton, E., et al. (2001). Investigation of human brain hemodynamics by simultaneous near-infrared spectroscopy and functional magnetic resonance imaging. *Medical Physics*, *28*, 521–527.

Treisman, A. M., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Bulletin*, *95*, 15–48.

Tulving, E., Kapur, S., Craik, F. I., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission topography findings. *Proceedings of the National Academy of Sciences USA*, *91*, 2016–2020.

Warm, J. S., & Dember, W. N. (1998). Tests of a vigilance taxonomy. In R. R. Hoffman, M. F. Sherick, & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 87–112). Washington, DC: American Psychological Association.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183–200). Mahwah, NJ: Erlbaum.

Warm, J. S., & Jerison, H. J. (1984). The psychophysics of vigilance. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 15–59). Chichester, UK: Wiley.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). New York: Academic Press.

Wickens, C. D. (1990). Applications of event-related potential research to problems in human factors. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson (Eds.), *Event-related brain potentials: Basic and applied issues* (pp. 301–309). New York: Oxford University Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*, 159–177.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Wiener, E. L., & Attwood, D. A. (1968). Training for vigilance: Combined cueing and knowledge of results. *Journal of Applied Psychology*, *52*, 474–479.

Prevention of Rear-End Crashes in Drivers with Task-Induced Fatigue through the Use of Auditory Collision Avoidance Warnings

Jennifer F. May and Carryl L. Baldwin
Old Dominion University, Department of Psychology
Raja Parasuraman
George Mason University, Department of Psychology

Driver fatigue is associated with increased lane position variability, decreased response time and increased crash rate. Collision avoidance systems (CASs), which are currently installed in many vehicles, have been shown to reduce the crash rates of alert drivers. The current investigation examined the effectiveness of two auditory CAS warnings for reducing crash rates in fatigued young and older drivers faced with a potential rear-end collision. Relative to a no-warning condition, auditory CAS warnings reduced overall crash probability. The presence of a CAS warning was particularly beneficial to reducing crashes in older drivers. Results of this study indicate that auditory CAS warnings may help reduce fatigue-related rear-end crashes, particularly among older drivers.

Fatigue and sleepiness account for approximately 56,000 crashes annually (Knipling & Wang, 1994). These statistics may underestimate the problem because unlike alcohol impairment, there are currently no standardized procedures for testing for or reporting fatigue, and as such crashes that were caused by fatigue may often be attributed to other factors, such as inattention. To further examine this issue, the present study examined the effectiveness of auditory collision avoidance system (CAS) warnings in preventing crashes among fatigued younger and older drivers.

Fatigue due to extended driving periods results in increased lane deviation, speed fluctuations and slowing of reaction time. Such performance decrements can be particularly evident in monotonous conditions such as nighttime or freeway driving. van der Hulst, Meijman & Rothengatter (2001), for example, showed that task-induced fatigue—1.25 hours of route memorization— increased subjective fatigue in a subsequent 30 minute monotonous driving session compared to a preceding 30 minute session.

Fatigue effects have also been observed after only 40 minutes of monotonous simulator driving, and the frequency of large steering movements (which translates into greater lane position variability) increased over the drive (Thiffault & Bergeron, 2003). Lenne, Triggs, and Redman (1998) found that both sleep deprivation and time of day led ot increased lane position variability over time in a simulated driving task.

Performing a secondary task while driving can also produce subjective fatigue and driving performance decrements during and post-fatigue induction. The present study examined this form of fatigue, referred to as an active form (Desmond & Hancock, 2001) of task-induced fatigue (Desmond & Matthews, 1997; Matthews & Desmond, 2002), in conjunction with time-on-task. Active, task-induced fatigue increases lane deviation and decreases perceptual sensitivity, particularly when drivers are traversing along straight roads (Desmond & Matthews, 1997; Matthews & Desmond, 2002).

CASs have the potential to mitigate crash risk stemming from active task-induced fatigue. CASs represent one of a variety of automation aids that can be implemented to support the human operator, in this case the driver. As such human factors evaluations of such advanced driver automation tools must be conducted so that their safety benefits can be fully realized (Hancock & Parasuraman, 1992). A combined visual-auditory CAS warning has been shown to reduce rear-end crashes among alert drivers in simulation studies (Lee, McGehee, Brown, & Reyes, 2002).

Additionally, verbal CAS warnings have been shown to reduce crash rates in a variety of high crash risk simulated driving scenarios. For example, Baldwin and May (2005) examined the effectiveness of verbal collision warnings as a function of signal word and presentation levels in reducing collisions among young drivers in simulated high-risk urban driving scenarios. They observed an acoustic-semantic tradeoff that significantly impacted crash rates. Specifically, the word "caution" at a high presentation level and "danger" at a low presentation level were effective in reducing crashes.

The current study was designed to examine the effectiveness of verbal and

nonverbal auditory CAS warnings in reducing crashes in fatigued older and younger drivers. Fatigue was induced by having participants perform a simulated driving task in combination with a secondary task for approximately 1.5 hours. When drivers demonstrated fatigue (as assessed by excessive lane position variability) a critical rear-end potential collision event was triggered. We predicted that driver performance would also be improved with the CAS warning. Drivers received either no warning or a verbal or nonverbal CAS warning.

We hypothesized that CAS condition, age and gender would significantly predict CAS response (crash or stopped). We also hypothesized that participants who stopped in time to avoid the collision would have a greater stopping distance if they received a CAS warning. Finally, of the participants who crashed, we hypothesized that participants who received a CAS warning would have a lower speed at the time of impact (presumably as a result of attempting to stop). Speed at time of impact could be taken as an index of crash severity in this simulation scenario.

METHOD

Participants

Forty-eight participants between 18 to 82 (M = 46.56, SD = 22.52); including 19 between the age of 18-35 (5 males and 14 females) and 23 between the ages of 60 and 82 (11 males and 12 females) completed the experiment. Three participants were excluded as outliers (defined as 3 SD above the mean) due to either their excessive following distance or speed deviation. All participants were licensed drivers and 95.6% of participants reported driving at least three to five times per week. A majority (93.3%) of participants reported using a computer and 31.1% reported playing video games. All participants received scores of over 26 on the mini-mental state exam (Folstein, Folstein, & McHugh, 1975), had normal hearing as assessed by a pure-tone audiometric exam and self-reported normal or corrected to normal vision.

Procedure

Participants were first familiarized with various CAS warnings and rated each for its perceived urgency, alerting effectiveness and annoyance level (Baldwin & May, 2005).

After a brief orientation to the driving simulator, participants completed a practice drive. A General Electric Capital I-SIM driving simulator was used to present the simulated driving task. The simulator consists of a built up cab unit with wrap around screens presenting an approximately 180 degree field of view. Participants use normal vehicle controls (i.e., steering wheel, accelerator, brake) to maintain their position and apparent speed on the simulated roadway.

The practice drive consisted of a car-following task, with traffic included in the second half of the drive. The lead car maintained a constant speed of 55 mph. This served to regulate the speed at which the participant could drive. Driving much faster than 55 mph would result in overtaking the lead car and driving much slower would result in losing the lead car. In this way, individual differences in driving speed were minimal. The practice drive also included a CAS warning ("notice") related to a car stranded on the side of the road which the participant had to maneuver around, so the participants would have experience with a CAS event.

Fatigue inducement. Participants then completed approximately 1.5 hours of simulated driving consisting of a car following task and secondary speech processing task in both high density and no traffic freeway scenarios. Included in this part of the experiment was a baseline car-following drive where average lane position variability (average standard deviation of their lane position, or amount of weaving) was calculated. The researchers used this average lane position variability and calculated one standard deviation above this average as an index of task-induced fatigue (fatigue threshold).

Collision scenario. The potential collision scenario consisted of a car following task with no other traffic. Lane position variability was monitored in real-time to determine fatigue level. Once the participant reached his or her fatigue index the researcher triggered the lead car to suddenly decelerate and come to a rapid complete stop. When the lead car slowed to 50 mph (from 55mph), three possible CAS conditions occurred. Sixteen participants had no warning (control condition), 15 participants heard the 1000 Hz tone and 14 participants received the verbal "danger" CAS

warning. Participants received no prior information regarding this potential crash situation but had been exposed to the CAS warnings during an earlier portion of the investigation.

The main variable of interest was response to the potential collision situation (crashed or stopped). Speed variability (standard deviation) and average following distance were also calculated. Stopping distance from lead car was recorded if the participant did not crash, and speed at moment of impact was recorded if the participant crashed.

RESULTS

Twenty-nine percent (13 of 45) of drivers crashed. Of these, 61.5% (8 of 13) crashes occurred in the no warning group, 23% (3 of 13) occurred in nonverbal warning group and 15% (2 of 13) occurred in the verbal warning group (see Figure 1). The distribution of crashes as a function of warning condition and age are graphically represented in Figure 2. Note that the age restrictions for older and younger groups resulted in the exclusion of 3 participants between the ages of 35 and 60, one of whom crashed in the no-warning condition. For all participants, the mean average following distance was 49.54 meters (*SD* = 19.67) and the mean speed deviation was calculated as 3.37 (*SD* = 1.32) mph.

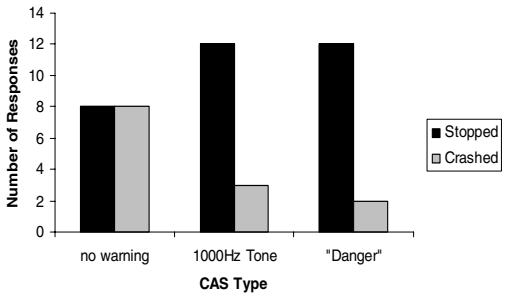


Figure 1. Number of Responses (Crashed or Stopped) by CAS Type.

A binary logistic step forward regression was performed to identify significant predictors of CAS response (stopped vs. crashed). Variables included in the analysis were CAS type (no warning, 1000 Hz tone, "Danger"), age, gender, average following distance, speed deviation and fatigue threshold. CAS type was the only significant predictor of CAS response (β = -.975, p = 0.037, Nagelkerke R^2 = .151).

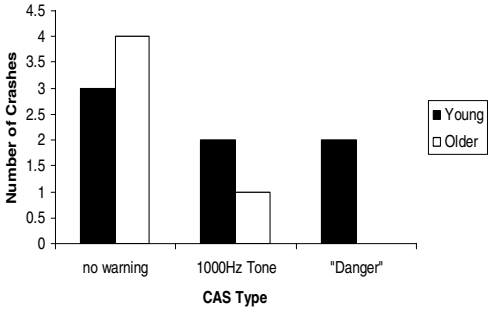


Figure 2. Number of Crashes by Age Group and CAS Type.

Separate regression analyses were run to examine the impact of CAS type in the young and older drivers. CAS type was a strong predictor of crash probability in the older group (β = -2.207, p = 0.05, Nagelkerke R^2 = .392), but was not a significant predictor for the younger group. No other variables were significant predictors of crash probability in either subgroup.

A 2 (age: young and older) by 3 (CAS type: nonverbal, verbal, control) MANOVA was performed to examine differences in CAS type and age on speed deviation, average following difference and fatigue threshold. The Wilk's Lambda criterion was used for this analysis. A significant multivariate main effect for age was revealed, $F(3, 34) = 4.81$, $p = .007$, $\eta^2 = .30$. Univariate analyses showed average following distance accounted for this age effect, $F(1, 36) = 13.57$, $p = .001$, $\eta^2 = .25$. Older drivers maintained a larger average following distance (M = 58.9, SD = 18.12) than the younger drivers (M = 39.3, SD = 17.67). No other significant main effects or interactions were found.

A separate MANOVA was performed to examine speed deviation, average following distance and fatigue threshold as a function of collision response type (stopped versus crashed). No significant effects were found.

Separate ANOVAs were performed to examine the effect of CAS type on stopping distance (for participants who stopped, N = 31) and speed at time of collision (for those who crashed, N = 13). There was no significant difference in the stopping distance between the three CAS types for participants who stopped,

nor was there a significant difference for speed at time of collision for participants who crashed. Again, these analyses were repeated to include age group. Results showed a significant age effect for stopping distance, $F(1, 23) = 6.6$, $p = .017$, $\eta^2 = .22$. Older drivers had a larger stopping distance ($M = 17.2$, $SD = 9.59$) than the younger drivers ($M = 9.81$, $SD = 4.22$). Speed at time of collision did not significantly differ between the two groups.

DISCUSSION

The results indicate that auditory CAS warnings are helpful in reducing crashes in high risk rear end collision situations when drivers are fatigued. Nearly 18% of drivers crashed when provided no prior warning. This observation confirms the success of our method of operationally defining a high collision risk situation in the simulation scenario used in the current investigation. When provided with an auditory warning, only roughly 11% of drivers crashed. This represents a significant reduction in crash rate relative to the no warning control condition. Auditory warnings were particularly effective in reducing collision potential among older drivers. In fact, when given an auditory warning only one driver over the age of 60 was unable to avoid collision. Unfortunately, providing an auditory warning had little impact on crash rates in drivers under the age of 35.

Average following distance, speed deviation and fatigue threshold did not differ as a function of CAS warning or between those who crashed and stopped. This finding is quite positive in that we know that following too close was not a cause for the crashes, and that fatigue levels appear equal across the groups. This was also evident in that these variables were not predictors of CAS response in the regression model.

Average following distance was larger for the drivers over the age of 60, indicating that these drivers may be more cautious when following a lead car, or that they increase their following distance more when fatigued, as compared to younger drivers. Previous literature has documented that fatigue was associated with increased headway during a monotonous driving task (van der Hulst et al., 2001). The increase in following distance coupled with an auditory collision warning may help older drivers avoid crashes. It is important to point out that despite the greater following distance maintained by older adults, when provided no prior auditory warning a substantial number (just over 17%) of these drivers crashed. Therefore the current results suggest that an auditory CAS warning may be particularly effective in reducing crash rates among older drivers experiencing task-induced fatigue.

For drivers who crashed, speed at time of impact did not vary across the CAS conditions. The potential for CAS warnings to reduce crash severity may have been obscured by low statistical power in the current investigation. Despite the relatively high rate of crashes (29%) across drivers a total of only 13 crashes were observed. Further examination of the potential for CAS warnings to reduce crash severity in a larger sample size is warranted. The significant reduction in crash rates among drivers provided with a warning indicate that auditory CASs have to potential to reduce both occurrence and severity of fatigue-related rear-end crashes, especially among older drivers.

In terms of the direct implications for task-induced fatigue-related crashes, a limitation of the current investigation is the exclusion of a control group exposed to the collision scenario and CAS conditions in the absence of fatigue inducement. In this way the effectiveness of the CAS warning could be compared in fatigued and non-fatigued drivers. In addition, subjective measures of fatigue could have supplemented the lane deviation measure to ensure the drivers were in fact fatigued.

The results point to the utility of auditory CAS warnings adaptively linked to measures of driver performance decrement. In the present study fatigue-related driving impairment was indexed by excessive lane position variability. After this fatigue inducement, warnings were presented in response to a high risk collision situation. Future research could examine a more prospective approach by investigating the potential benefit of providing drivers with low hazard level visual or auditory warnings when driving performance degraded irrespective of a potential collision situation. The auditory CAS warnings examined in this study could also be presented adaptively, based on individual driver performance characteristics (e.g., excessive fatigue-related lane deviation) or aspects of the driving environment, given that adaptive automation has been shown to regulate

operator workload and performance (Parasuraman, 2000). Future research on adaptive automation to mitigate fatigue in drivers might also consider other performance and physiological measures, including EEG, blink rate, or eye movements.

ACKNOWLEDGEMENTS

This work was funded in part by a grant to R.P. from the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office (DAAD19-01-1-0621). We would also like to acknowledge Ian Reagan and Martin Smith-Rodden for their help in the collection of this data.

References

Baldwin, C. L., & May, J. F. (2005, June). *Verbal Collision Avoidance Messages of Varying Perceived Urgency Reduce Crashes In High Risk Scenarios.* Paper presented at the Driving Assessment 2005, Rockport, Maine.

Desmond, P. A., & Hancock, P. A. (2001). Active and passive fatigue stress. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 455-465). Mahwah, NJ: Lawrence Erlbaum.

Desmond, P. A., & Matthews, G. (1997). Implications of task-induced fatigue effects for in-vehicle countermeasures to driver fatigue. *Accident Analysis & Prevention, 29*(4), 515-523.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). 'Mini-Mental-State'. A practical method for grading the cognitive state of patients for the clinicians. *Journal of Psychiatric Research, 12*, 189-198.

Hancock, P. A., & Parasuraman, R. (1992). Human factors and safety issues in Intelligent Vehicle Highway Systems (IVHS). *Journal of Safety Research, 23,* 181-198.

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44*(2), 314-334.

Lenne, M.G. Triggs, T.J. & Redman, J.R. (1998). Interactive effects of sleep deprivation, time of day and driving experience on a driving task. *Sleep, 21(1)*, 38-44.

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 55*(2), 659-686.

Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics, 43*, 931-951.

Rosen, I. M. (2005). Driving while sleepy should be a criminal offense. *Journal of Clinical Sleep Medicine, 1*(4), 337-340.

van der Hulst, M., Meijman, T. & Rothengatter, T. (2001). Maintaining task set under fatigue: a study of time-on-task effects in simulated driving. *Transportation Research Part F, 4*, 103-118.

Thiffault, P. & Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis and Prevention, 35*, 381-391.

Interim Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield
Army MURI, Subcontract to U. Central Florida

Raja Parasuraman
George Mason University

PERFORMANCE PERIOD: 5/1/05-7/30/05

The following study was completed during this performance period.

An Examination of the Relationship Between Cerebral Blood Flow and Single and Multitask Performance at Varying Levels of Difficulty

Adaptive automation has been proposed as a way of mitigating stress, workload, and fatigue effects by real-time monitoring of the operator. However, real-time measures of mental workload for use in adaptive automation studies must be able to distinguish between sources of workload. Two such sources are the number of concurrent tasks performed and the difficulty of each task. The present study examined both these sources using Transcranial Doppler Sonography (TCD) to measure cerebral blood flow velocity (BFV).

A total of 12 right-handed subjects (9 female) were run (age range 18-24, M=19.84). Data from 2 subjects (both male) was thrown out because of inconsistent BFV signal. One subject could only be isonated using the right middle cerebral artery (MCA). The group consisted of male and female college students who participated as part of a requirement of an introductory psychology course at George Mason University. The Edinburgh Handedness Inventory (Oldfield, 1970), the NASA Task Load Index (NASA-TLX), and the Dundee Stress State Questionnaire (DSSQ; Matthews, Joyner, Gilliland, Campbell, Falconer & Huggins, 1999) were administered to each participant. Participants performed the Synwork (Elsmore, 1994) computer-based task battery. Synwork consists of four tasks: Sternberg Memory, 3-Column Addition, Visual Tracking and Auditory Monitoring, of which only the first two tasks were, Sternberg Memory and Addition (Math) were used. Three levels of task were manipulated: task type (memory and math), task difficulty (easy and difficult) and number of concurrent tasks (single and multi). The conditions were counterbalanced across task type and difficulty but not across number of tasks. The TCD Unit by Nicolet Vascular (Model: Pioneer TC8080), was used to measure BFV through the transtemporal window to directly assess the middle cerebral artery. The probes were stabilized and held in place by a headset.

As of the date of this progress report, data collection has been completed and the results are being analyzed.

Interim Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield
Army MURI, Subcontract to U. Central Florida

Raja Parasuraman
George Mason University

PERFORMANCE PERIOD: 8/1/05-10/31/05

The following study conducted by research assistant Jennifer May, was started during this performance period. Data analysis is ongoing.

Evaluation of an Adaptive Collision Avoidance System for Fatigue Mitigation

Sixteen participants were tested on a GE driving simulator. Following performance of driving scenarios unrelated to this study, participants drove on a simulated highway for a continuous period of 30 minutes. Lane position variation (LPV), speed, and other driving parameters were continuously monitored. Brain electrical activity was also monitored using an electrode cap and a 32-channel EEG system. A moving-window estimate of lane position deviation of a participant driver was compared to a baseline established for that participant. Towards the end of the continuous drive, an unexpected event (a stopped vehicle) was presented. An auditory collision warning was sounded depending upon on the degree of deviation of lane position from baseline (LPV > 1 standard deviation of baseline). There were three conditions: no collision warning, verbal warning of "Danger" and a 1000 hz tone. Driving response parameters following the unexpected event and EEG parameters were analyzed. Data analysis is currently ongoing, The results are expected to have implications for the evaluation and design of physiologically-based adaptive automation for mitigating fatigue.

Interim Progress Report

Mitigating Stress, Workload, and Fatigue on the Electronic Battlefield
Army MURI, Subcontract to U. Central Florida

Raja Parasuraman
George Mason University

PERFORMANCE PERIOD: 1/1/06 – 8/31/06

A study on the effects of an adaptive collision warning system for mitigating task-related fatigue during prolonged simulated driving was conducted. Subject testing was continued but not completed while we were awaiting the last incremental funding from the University of Central Florida, which was considerably delayed in delivery from the Army sponsor.

Subject testing has now recommenced and is anticipated to be finished in October 2006.

Initial data analysis of the study participants to date has been carried out. The preliminary results will be presented at the Annual Meeting of the Human Factors and Ergonomics society.

May, J., Baldwin, C. L., & Parasuraman, R. (2006, October). Prevention of rear-end crashes in drivers with task-induced fatigue through the use of auditory collision avoidance warnings. To appear in *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society,* San Francisco, CA.

STANDARD

NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION

AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY
RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS
(MURI / OPUS) LABORATORY

AGREEMENT TITLE: Operator Under Stress Research

AGREEMENT NUMBER: NCRADA-**NAVAIR ORLANDO TSD-02-037**

AGREEMENT ADMINISTRATORS:

NAVAIR ORLANDO TSD

Technology Transfer Office:	**Ms. Patricia R. Nimmo, AIR 49T, (407) 380-8225**
Legal Counsel:	**Robert W. Adams, AIR 772P, (407) 380-8211**
Principal Investigator:	**Ron Wolff, AIR 4962, (407) 380-4583**

UCF MURI / OPUS

Preferred Contact:	**Mrs. Kim Smith, UCF Office of Research, (407) 823-3062**
Principal Investigator:	**Dr. Peter Hancock, Director, UCF MURI/OPUS Laboratory, (407) 823-2310**

This Page Intentionally Left Blank

TABLE OF CONTENTS

PREAMBLE

Article 1.

DEFINITIONS 2

1.1 "Agreement" 2
1.2 "Classified Information" 2
1.3 "Collaborator" 2
1.4 "Controlled Unclassified Information (CUI)" 2
1.5 "Cooperative Work" 2
1.6 "Data" 2
1.7 "Effective Date" 2
1.8 "Exclusive License" 2
1.9 "Government" 2
1.10 "Government Purpose Rights" 2
1.11 "Information" 2
1.12 "Intellectual Property" 2
1.13 "Invention" 2
1.14 "Invention Disclosure" 3
1.15 "Made" 3
1.16 "Militarily Critical Technologies (MCT)" 3
1.17 "Non-Subject Data" 3
1.18 "Non-Subject Invention" 3
1.19 "Patent Application" 3
1.20 "Principal Investigator (PI)" 3
1.21 "Proprietary Information" 3
1.22 "Restricted Access Information" 3
1.23 "Subject Data" 3
1.24 "Subject Invention" 3
1.25 "Tangible Property" 4
1.26 "Unlimited Rights" 4

Article 2. OBJECTIVES 4

Article 3. RESPONSIBILITIES 4

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities 4
3.2 **UCF MURI / OPUS Laboratory** Personnel and facilities 5
3.3 Security Regulations and Directives 5

Article 4. REPRESENTATIONS AND WARRANTIES 5

4.1 **NAVAIR ORLANDO TSD** Representations and Warranties 5
4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Representations and Warranties 5
4.3 Joint Representations 6

Article 5. FUNDING 7

5.1 Payment Schedule..... 7
5.2 Insufficient and Excess Funds..... 7
5.3 No New Commitments..... 8
5.4 Accounting Records..... 8

Article 6. REPORTS AND PUBLICATIONS..... 8

6.1 Final Reports..... 8
6.2 Agreement to Confer Prior to Publication or Public Disclosure..... 9
6.3 Classified Information..... 9

Article 7. INTELLECTUAL PROPERTY..... 9

7.1 Data..... 9
7.1.1 General Provisions Applying to All Data..... 9
7.1.1.1 Ownership..... 9
7.1.1.2 No Implied License..... 9
7.1.1.3 Marking of Data..... 10
7.1.1.3.1 Data Provided With Less Than Unlimited Rights.. 10
7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information..... 10
7.1.1.3.3 Data That are Subject to 35 USC 205..... 11
7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted..... 11
7.1.1.4 Protection of Data..... 11
7.1.1.5 Release of Data Under the Freedom of Information Act..... 11
7.1.2 Subject Data..... 12
7.1.2.1 Delivery of Requested Subject Data..... 12
7.1.2.2 Rights in Subject Data..... 12
7.1.3 Rights in Non-Subject Data..... 12
7.2 Copyrights..... 13
7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory**..... 13
7.2.2 Copyright License to the Government..... 13
7.2.3 Copyright Statement..... 13
7.3 Trademarks and Service Marks..... 13
7.3.1 Ownership of Trademarks and Service Marks..... 13
7.3.2 Obligation of Employees to Report Trademarks and Service Marks.... 13
7.3.3 Obligation of Collaborators to Notify Each Other..... 13
7.3.4 Responsibility for Filing an Application for Trademark or Service Mark..... 14
7.3.5 License to Use Trademark or Service Mark..... 14
7.4 Subject Inventions..... 14
7.4.1 Obligation to Report Subject Inventions..... 14
7.4.1.1 Collaborators' Instruction to Employees..... 14
7.4.1.2 Timely Invention Disclosure by Inventors..... 14
7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator..... 14
7.4.2 Determination of Subject Inventions..... 14
7.4.3 Title to and Ownership of Subject Inventions..... 14

7.4.4 Filing of Patent Applications..... 15
7.4.4.1 Filing of Patent Applications on Solely Made Inventions..... 15
7.4.4.2 Filing of Patent Applications on Jointly Made Inventions..... 15
7.4.4.3 Preserving Intellectual Property Rights..... 15
7.4.4.4 Filing Deadlines..... 15
7.4.4.5 Copies and Inspection..... 15
7.4.4.5.1 Copies of Prosecution Papers..... 15
7.4.4.5.2 Access to Patent Application File and Right to Make Copies..... 16
7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application..... 16
7.4.5 Nonexclusive License to Subject Inventions..... 16
7.4.5.1 Nonexclusive License Grant..... 16
7.4.5.2 Confirmatory Nonexclusive License Agreement..... 16
7.4.6 Option for Exclusive License to Subject Inventions..... 16
7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement. 17
7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions..... 17
7.4.8.1 Exclusive Licenses and Exclusive License Option..... 17
7.4.8.2 Nonexclusive Licenses..... 17
7.5 Non-Subject Inventions..... 18
7.5.1 Ownership of Non-Subject Inventions..... 18
7.5.2 Rights Under Other Agreements..... 18
7.5.3 No License to Non-Subject Inventions..... 18
7.6 Research License..... 18

Article 8. TANGIBLE PROPERTY..... 18

8.1 Title to Preexisting Tangible Property..... 18
8.2 Tangible Property Purchased by Collaborators to Perform Cooperative Work... 18
8.3 Title to Developed Tangible Property..... 19
8.4 Tangible Property Operational and Disposition Costs..... 19
8.5 Disposal of Tangible Property..... 19

Article 9. LIABILITY..... 19

9.1 Extent of Government Liability..... 19
9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of The UCF MURI / OPUS Laboratory** Liability..... 19
9.3 *Force Majeure*..... 20

Article 10. GENERAL PROVISIONS..... 20

10.1 Characteristics of the Agreement..... 20
10.1.1 Entire Agreement..... 20
10.1.2 Severability..... 20
10.1.3 Interpretation of Headings..... 20
10.2 Agreement Between Collaborators..... 20
10.2.1 Governing Laws..... 20
10.2.2 Independent Parties/Entities..... 20
10.2.3 Assignment /Subcontracting..... 20

10.2.4 Disputes..... 21
10.2.4.1 Settlement and Resolution..... 21
10.2.4.2 Continuation of Cooperative Work..... 21
10.2.5 Waivers..... 21
10.2.6 Use of Name or Endorsements..... 22
10.3 Environment, Safety, and Health..... 22
10.4 U.S. Competitiveness..... 22
10.5 Public Release of This Agreement..... 22

Article 11. MODIFICATIONS AND NOTICES..... 22

11.1 Amendments..... 22
11.2 Termination..... 23
11.2.1 Termination by Mutual Consent..... 23
11.2.2 Unilateral Termination..... 23
11.3 Notices..... 23

Article 12. SURVIVING PROVISIONS..... 24

Article 13. DURATION..... 24

Article 14. SIGNATURES..... 24

Appendix A
Statement of Work

Appendix B
Confirmatory License Agreement

STANDARD
NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY
RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS
(MURI / OPUS) LABORATORY

PREAMBLE

Under authority of the U.S. Federal Technology Transfer Act of 1986 (Public Law 99-502, 20 October 1986, as amended), **NAVAIR ORLANDO TSD (Naval Air Warfare Center Training Systems Division)**, located at **12350 Research Parkway, Orlando, Florida 32826-3275**, and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, whose administrative offices are located at **12443 Research Parkway, Suite 207, Orlando, FL 32826-3252**, enter into this Cooperative Research and Development Agreement (CRADA), which shall be binding upon the Collaborators and their assignees according to the clauses and conditions hereof and for the term and duration set forth.

The U.S. Federal Technology Transfer Act of 1986, as amended, provides for making the expertise, capabilities, and technologies of U.S. Federal laboratories accessible to other Federal agencies; units of State or local government; industrial organizations (including corporations, partnerships and limited partnerships, and industrial development organizations); public and private foundations; nonprofit organizations (including universities); or other persons in order to improve the economic, environmental, and social well-being of the United States by stimulating utilization of U.S. Federally funded technology developments and/or capabilities.

NAVAIR ORLANDO TSD has extensive expertise, capabilities, and information in **Virtual Environments for Operator Training**, and, in accordance with the U.S. Federal Technology Transfer Act, desires to make this expertise and technology available for use in the public and private sectors.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory has the interest, resources, capabilities, and technical expertise to transition the results of Naval research and development for public use.

NOW THEREFORE, the Collaborators agree as follows.

Article 1. DEFINITIONS

As used in this Agreement, the following terms shall have the meanings defined below, which are equally applicable to both the singular and plural forms of nouns or any tense of verbs.

1.1 "Agreement" means this Cooperative Research and Development Agreement (CRADA) with its Appendices.

1.2 "Classified Information" means all Data classified in accordance with the national security laws of the United States.

1.3 "Collaborator" means the Navy participant or the Non-Navy participant represented and bound by the signatories of this Agreement.

1.4 "Controlled Unclassified Information (CUI)" means Government Data, Information, or materials provided to or resulting from this Agreement that may be export controlled, sensitive, for official use only, or otherwise protected by law, executive order, or regulation.

1.5 "Cooperative Work" means research, development, engineering, or other tasks performed under this Agreement by **NAVAIR ORLANDO TSD** or **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** working individually or together, pursuant to the Objectives (Article 2) and the Statement of Work (Appendix A).

1.6 "Data" means recorded information of any kind regardless of the form or method of the recording, including computer software.

1.7 "Effective Date" means the date of the last signature of the Collaborators executing this Agreement.

1.8 "Exclusive License" means the grant by the owner of Intellectual Property of the exclusive right to make, use, or sell a patented invention.

1.9 "Government" means the Government of the United States of America.

1.10 "Government Purpose Rights" means the right of the Government to use, duplicate, or disclose Data, in whole or in part, and in any manner, for Government purposes only, and to have or permit others to do so for Government purposes only. Government Purpose Rights includes competitive procurement, but does not include the right to have or permit others to use Data for commercial purposes.

1.11 "Information" means all data, trade secrets, and commercial and financial information. (Chapter 5 Subsection II of Title 5 USC)

1.12 "Intellectual Property" means the property of ideas, examples of which include, but are not limited to, patents, trademarks, copyrights, and trade secrets.

1.13 "Invention" means any invention or discovery that is or may be patentable or otherwise protected under Title 35, United States Code, or any novel variety of plant that is or may be patentable under the Plant Variety Protection Act. (15 USC 3703(9))

1.14 "Invention Disclosure" means the document identifying and describing to organizational management the Making of an Invention.

1.15 "Made" when used in conjunction with any Invention means the conception or first actual reduction to practice of such Invention. (15 USC 3703(10))

1.16 "Militarily Critical Technologies (MCT)" means those technologies identified in the Militarily Critical Technologies List and under the Export Administration Act of 1979, as amended.

1.17 "Non-Subject Data" means any Data that are not Subject Data.

1.18 "Non-Subject Invention" means any Invention that is not a Subject Invention.

1.19 "Patent Application" means an application for patent protection for an Invention with any domestic or foreign patent-issuing authority.

1.20 "Principal Investigator (PI)" means that person having the responsibility for the performance of the Cooperative Work on behalf of a Collaborator.

1.21 "Proprietary Information" means information that embodies trade secrets developed at private expense or business, commercial, or financial information that is privileged or confidential provided that such information:

is not known or available from other sources without obligations concerning its confidentiality;

has not been made available by the owners to others without obligation concerning its confidentiality;

is not already available to the Government without obligation concerning its confidentiality; and

has not been developed independently by persons who have had no access to the information. (FAR/DFARS Definition)

1.22 "Restricted Access Information" means Subject Data generated by **NAVAIR ORLANDO TSD** that would be Proprietary Information if the Information had been obtained from a non-Federal Collaborator participating in a CRADA (15 USC 3710a). Under 15 USC 3710a(c)(7)(B), the Collaborators mutually may agree to provide appropriate protection to Subject Data generated by **NAVAIR ORLANDO TSD** (Restricted Access Information) against public dissemination or release under the Freedom of Information Act (FOIA) for a period of up to five (5) years after development of the Information.

1.23 "Subject Data" means that Data first recorded in the performance of the Cooperative Work.

1.24 "Subject Invention" means any Invention Made in the performance of the Cooperative Work.

1.25 "Tangible Property" means personal or real property that can be physically touched or held.

1.26 "Unlimited Rights" means the right to use, modify, reproduce, release, disclose, perform, or display Data or Computer Programs in whole or in part, in any manner and for any purpose whatsoever, and to have or permit others to do so.

Article 2. OBJECTIVES

Background

NAVAIR ORLANDO TSD has been conducting Navy sponsored research in virtual environments for operator/small arms training for over 15 years. Results of this research have already been transitioned for many military training applications, and have also been the basis for technology transition in use by law enforcement. The Small Arms Simulation Training (SAST) laboratory at **NAVAIR ORLANDO TSD** now has in use an extensive range of software and data bases that have resulted from in-house **NAVAIR ORLANDO TSD** research work.

The University of Central Florida (UCF), with plans to establish a new laboratory for Human Factors research related to Homeland Defense, became aware of the work already done in **NAVAIR ORLANDO TSD's** SAST laboratory. Discussions were immediately begun to explore the mutual benefits of an agreement to work collaboratively.

The primary objective of this work will be to enhance mutual understand of operator performance under stress and develop possible applications of human factors research in this area for military training, and for homeland defense and related technologies

Benefit to Navy Partner: Research using **NAVAIR ORLANDO TSD** developed or enhanced source code may be expected to result in further enhancements to software and databases, at no expense to the government, and with direct use in military simulation-based training. Additionally, the anticipated UCF human factors research projects may be expected to have direct application and benefit for military training. Further, expertise will be provided to UCF on a cost basis, leveraging Navy research dollars.

Benefit to Non-Navy partner: **NAVAIR ORLANDO TSD** developed or enhanced source code for scenario generation, rural and urban databases, weapon tracking, image generation, networking/communication, etc., can provide the kind of virtual environment test bed required for planned human factors research, and can greatly accelerate work toward research objectives.

Article 3. RESPONSIBILITIES

The Collaborators shall provide personnel, facilities, and equipment necessary for, and shall perform, the Cooperative Work.

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities

The Cooperative Work done by **NAVAIR ORLANDO TSD** will be performed under the program guidance of **Ron Wolff**, PI, **NAVAIR ORLANDO TSD, AIR 4962**, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed

within the facilities of **NAVAIR ORLANDO TSD** or done on behalf of **NAVAIR ORLANDO TSD** by third parties in support of this Agreement.

UCF MURI / OPUS Laboratory personnel who perform Cooperative Work at **NAVAIR ORLANDO TSD** facilities will be supervised by the **UCF MURI / OPUS Laboratory** PI.

3.2 **UCF MURI / OPUS Laboratory** Personnel and Facilities

The Cooperative Work done by the **UCF MURI / OPUS Laboratory** will be performed under the program guidance of **Dr. Peter Hancock, UCF MURI / OPUS Laboratory** PI, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed within the facilities of the **UCF MURI / OPUS Laboratory** or done on behalf of the **UCF MURI / OPUS Laboratory** by third parties in support of this Agreement.

NAVAIR ORLANDO TSD personnel who perform Cooperative Work at the **UCF MURI / OPUS Laboratory** facilities will be supervised by **NAVAIR ORLANDO TSD** PI.

3.3 Security Regulations and Directives

Each Collaborator will abide by the safety and security regulations and directives of the host facility in which the Cooperative Work is being performed.

Article 4. REPRESENTATIONS AND WARRANTIES

4.1 **NAVAIR ORLANDO TSD**'s Representations and Warranties

NAVAIR ORLANDO TSD hereby warrants and represents to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** as follows:

4.1.1 **NAVAIR ORLANDO TSD** is a Federal laboratory of the U.S. Department of the Navy (Navy) as defined by 15 USC 3710a(d)(2)(A) and Department of Defense Instruction 5535.8, dated May 14, 1999.

4.1.2 The performance of the activities specified by this Agreement is consistent with the **Virtual Environments for Operator/Small Arms Training** and technology transfer missions of **NAVAIR ORLANDO TSD** (15 USC 3710a).

4.1.3 The Department of the Navy official executing this Agreement for **NAVAIR ORLANDO TSD** has the requisite power and authority to enter into this Agreement and to bind **NAVAIR ORLANDO TSD** to perform according to the terms of this Agreement.

4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Representations and Warranties

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory hereby warrants and represents to **NAVAIR ORLANDO TSD** as follows:

4.2.1 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not directly or indirectly controlled by a foreign company or government (Executive Order 12591, Section 4 (a)). **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, as of the Effective Date of this Agreement, is a part of a **university** duly organized, validly existing, and in good standing under the laws of **the State of Florida.**

4.2.2 The official executing this Agreement for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** has the requisite power and authority to enter into this Agreement and to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to perform according to the terms of this Agreement.

4.2.3 The University of Central Florida and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** have taken all actions required by law, or otherwise, to authorize the execution and delivery of agreements, such as this Agreement.

4.2.4 The execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any agreement binding on **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. Furthermore, the execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any valid order of any court, or any regulatory agency or other body having authority to which **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is subject.

4.2.5 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not presently subject to debarment or suspension by any agency of the Government. Should **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** be debarred or suspended during the term of this Agreement or thereafter, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** will notify **NAVAIR ORLANDO TSD** within thirty (30) days of receipt of a final notice. **NAVAIR ORLANDO TSD** may then elect to terminate this Agreement and any licenses and options granted under this Agreement.

4.2.6 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is not** a small business as defined in 15 USC 632 and implementing regulations (13 CFR 121.101 et seq.) of the Administrator of the Small Business Administration.

4.3 Joint Representations

The Collaborators make the following representations.

STANDARD
NAVY COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT

BETWEEN

NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND

THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH,
UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY
RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS
(MURI / OPUS) LABORATORY

PREAMBLE

Under authority of the U.S. Federal Technology Transfer Act of 1986 (Public Law 99-502, 20 October 1986, as amended), **NAVAIR ORLANDO TSD (Naval Air Warfare Center Training Systems Division)**, located at **12350 Research Parkway, Orlando, Florida 32826-3275**, and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, whose administrative offices are located at **12443 Research Parkway, Suite 207, Orlando, FL 32826-3252**, enter into this Cooperative Research and Development Agreement (CRADA), which shall be binding upon the Collaborators and their assignees according to the clauses and conditions hereof and for the term and duration set forth.

The U.S. Federal Technology Transfer Act of 1986, as amended, provides for making the expertise, capabilities, and technologies of U.S. Federal laboratories accessible to other Federal agencies; units of State or local government; industrial organizations (including corporations, partnerships and limited partnerships, and industrial development organizations); public and private foundations; nonprofit organizations (including universities); or other persons in order to improve the economic, environmental, and social well-being of the United States by stimulating utilization of U.S. Federally funded technology developments and/or capabilities.

NAVAIR ORLANDO TSD has extensive expertise, capabilities, and information in **Virtual Environments for Operator Training**, and, in accordance with the U.S. Federal Technology Transfer Act, desires to make this expertise and technology available for use in the public and private sectors.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory has the interest, resources, capabilities, and technical expertise to transition the results of Naval research and development for public use.

NOW THEREFORE, the Collaborators agree as follows.

Article 1. DEFINITIONS

As used in this Agreement, the following terms shall have the meanings defined below, which are equally applicable to both the singular and plural forms of nouns or any tense of verbs.

1.1 "Agreement" means this Cooperative Research and Development Agreement (CRADA) with its Appendices.

1.2 "Classified Information" means all Data classified in accordance with the national security laws of the United States.

1.3 "Collaborator" means the Navy participant or the Non-Navy participant represented and bound by the signatories of this Agreement.

1.4 "Controlled Unclassified Information (CUI)" means Government Data, Information, or materials provided to or resulting from this Agreement that may be export controlled, sensitive, for official use only, or otherwise protected by law, executive order, or regulation.

1.5 "Cooperative Work" means research, development, engineering, or other tasks performed under this Agreement by **NAVAIR ORLANDO TSD** or **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** working individually or together, pursuant to the Objectives (Article 2) and the Statement of Work (Appendix A).

1.6 "Data" means recorded information of any kind regardless of the form or method of the recording, including computer software.

1.7 "Effective Date" means the date of the last signature of the Collaborators executing this Agreement.

1.8 "Exclusive License" means the grant by the owner of Intellectual Property of the exclusive right to make, use, or sell a patented invention.

1.9 "Government" means the Government of the United States of America.

1.10 "Government Purpose Rights" means the right of the Government to use, duplicate, or disclose Data, in whole or in part, and in any manner, for Government purposes only, and to have or permit others to do so for Government purposes only. Government Purpose Rights includes competitive procurement, but does not include the right to have or permit others to use Data for commercial purposes.

1.11 "Information" means all data, trade secrets, and commercial and financial information. (Chapter 5 Subsection II of Title 5 USC)

1.12 "Intellectual Property" means the property of ideas, examples of which include, but are not limited to, patents, trademarks, copyrights, and trade secrets.

1.13 "Invention" means any invention or discovery that is or may be patentable or otherwise protected under Title 35, United States Code, or any novel variety of plant that is or may be patentable under the Plant Variety Protection Act. (15 USC 3703(9))

1.14 "Invention Disclosure" means the document identifying and describing to organizational management the Making of an Invention.

1.15 "Made" when used in conjunction with any Invention means the conception or first actual reduction to practice of such Invention. (15 USC 3703(10))

1.16 "Militarily Critical Technologies (MCT)" means those technologies identified in the Militarily Critical Technologies List and under the Export Administration Act of 1979, as amended.

1.17 "Non-Subject Data" means any Data that are not Subject Data.

1.18 "Non-Subject Invention" means any Invention that is not a Subject Invention.

1.19 "Patent Application" means an application for patent protection for an Invention with any domestic or foreign patent-issuing authority.

1.20 "Principal Investigator (PI)" means that person having the responsibility for the performance of the Cooperative Work on behalf of a Collaborator.

1.21 "Proprietary Information" means information that embodies trade secrets developed at private expense or business, commercial, or financial information that is privileged or confidential provided that such information:

is not known or available from other sources without obligations concerning its confidentiality;

has not been made available by the owners to others without obligation concerning its confidentiality;

is not already available to the Government without obligation concerning its confidentiality; and

has not been developed independently by persons who have had no access to the information. (FAR/DFARS Definition)

1.22 "Restricted Access Information" means Subject Data generated by **NAVAIR ORLANDO TSD** that would be Proprietary Information if the Information had been obtained from a non-Federal Collaborator participating in a CRADA (15 USC 3710a). Under 15 USC 3710a(c)(7)(B), the Collaborators mutually may agree to provide appropriate protection to Subject Data generated by **NAVAIR ORLANDO TSD** (Restricted Access Information) against public dissemination or release under the Freedom of Information Act (FOIA) for a period of up to five (5) years after development of the Information.

1.23 "Subject Data" means that Data first recorded in the performance of the Cooperative Work.

1.24 "Subject Invention" means any Invention Made in the performance of the Cooperative Work.

1.25 "Tangible Property" means personal or real property that can be physically touched or held.

1.26 "Unlimited Rights" means the right to use, modify, reproduce, release, disclose, perform, or display Data or Computer Programs in whole or in part, in any manner and for any purpose whatsoever, and to have or permit others to do so.

Article 2. OBJECTIVES

Background

NAVAIR ORLANDO TSD has been conducting Navy sponsored research in virtual environments for operator/small arms training for over 15 years. Results of this research have already been transitioned for many military training applications, and have also been the basis for technology transition in use by law enforcement. The Small Arms Simulation Training (SAST) laboratory at **NAVAIR ORLANDO TSD** now has in use an extensive range of software and data bases that have resulted from in-house **NAVAIR ORLANDO TSD** research work.

The University of Central Florida (UCF), with plans to establish a new laboratory for Human Factors research related to Homeland Defense, became aware of the work already done in **NAVAIR ORLANDO TSD's** SAST laboratory. Discussions were immediately begun to explore the mutual benefits of an agreement to work collaboratively.

The primary objective of this work will be to enhance mutual understand of operator performance under stress and develop possible applications of human factors research in this area for military training, and for homeland defense and related technologies

Benefit to Navy Partner: Research using **NAVAIR ORLANDO TSD** developed or enhanced source code may be expected to result in further enhancements to software and databases, at no expense to the government, and with direct use in military simulation-based training. Additionally, the anticipated UCF human factors research projects may be expected to have direct application and benefit for military training. Further, expertise will be provided to UCF on a cost basis, leveraging Navy research dollars.

Benefit to Non-Navy partner: **NAVAIR ORLANDO TSD** developed or enhanced source code for scenario generation, rural and urban databases, weapon tracking, image generation, networking/communication, etc., can provide the kind of virtual environment test bed required for planned human factors research, and can greatly accelerate work toward research objectives.

Article 3. RESPONSIBILITIES

The Collaborators shall provide personnel, facilities, and equipment necessary for, and shall perform, the Cooperative Work.

3.1 **NAVAIR ORLANDO TSD** Personnel and Facilities

The Cooperative Work done by **NAVAIR ORLANDO TSD** will be performed under the program guidance of **Ron Wolff,** PI, **NAVAIR ORLANDO TSD, AIR 4962**, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed

within the facilities of **NAVAIR ORLANDO TSD** or done on behalf of **NAVAIR ORLANDO TSD** by third parties in support of this Agreement.

UCF MURI / OPUS Laboratory personnel who perform Cooperative Work at **NAVAIR ORLANDO TSD** facilities will be supervised by the **UCF MURI / OPUS Laboratory** PI.

3.2 **UCF MURI / OPUS Laboratory** Personnel and Facilities

The Cooperative Work done by the **UCF MURI / OPUS Laboratory** will be performed under the program guidance of **Dr. Peter Hancock, UCF MURI / OPUS Laboratory** PI, who has the responsibility for the scientific and technical conduct of the Cooperative Work performed within the facilities of the **UCF MURI / OPUS Laboratory** or done on behalf of the **UCF MURI / OPUS Laboratory** by third parties in support of this Agreement.

NAVAIR ORLANDO TSD personnel who perform Cooperative Work at the **UCF MURI / OPUS Laboratory** facilities will be supervised by **NAVAIR ORLANDO TSD** PI.

3.3 Security Regulations and Directives

Each Collaborator will abide by the safety and security regulations and directives of the host facility in which the Cooperative Work is being performed.

Article 4. REPRESENTATIONS AND WARRANTIES

4.1 **NAVAIR ORLANDO TSD**'s Representations and Warranties

NAVAIR ORLANDO TSD hereby warrants and represents to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** as follows:

4.1.1 **NAVAIR ORLANDO TSD** is a Federal laboratory of the U.S. Department of the Navy (Navy) as defined by 15 USC 3710a(d)(2)(A) and Department of Defense Instruction 5535.8, dated May 14, 1999.

4.1.2 The performance of the activities specified by this Agreement is consistent with the **Virtual Environments for Operator/Small Arms Training** and technology transfer missions of **NAVAIR ORLANDO TSD** (15 USC 3710a).

4.1.3 The Department of the Navy official executing this Agreement for **NAVAIR ORLANDO TSD** has the requisite power and authority to enter into this Agreement and to bind **NAVAIR ORLANDO TSD** to perform according to the terms of this Agreement.

4.2 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Representations and Warranties

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory hereby warrants and represents to **NAVAIR ORLANDO TSD** as follows:

4.2.1 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not directly or indirectly controlled by a foreign company or government (Executive Order 12591, Section 4 (a)). **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, as of the Effective Date of this Agreement, is a part of a **university** duly organized, validly existing, and in good standing under the laws of **the State of Florida.**

4.2.2 The official executing this Agreement for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** has the requisite power and authority to enter into this Agreement and to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to perform according to the terms of this Agreement.

4.2.3 The University of Central Florida and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** have taken all actions required by law, or otherwise, to authorize the execution and delivery of agreements, such as this Agreement.

4.2.4 The execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any agreement binding on **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**. Furthermore, the execution and delivery of this Agreement does not contravene any material provision of, or constitute a material default under, any valid order of any court, or any regulatory agency or other body having authority to which **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is subject.

4.2.5 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is not presently subject to debarment or suspension by any agency of the Government. Should **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** be debarred or suspended during the term of this Agreement or thereafter, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** will notify **NAVAIR ORLANDO TSD** within thirty (30) days of receipt of a final notice. **NAVAIR ORLANDO TSD** may then elect to terminate this Agreement and any licenses and options granted under this Agreement.

4.2.6 **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is not** a small business as defined in 15 USC 632 and implementing regulations (13 CFR 121.101 et seq.) of the Administrator of the Small Business Administration.

4.3 Joint Representations

The Collaborators make the following representations.

4.3.1 There is no express or implied warranty as to any research, Invention, or product, whether tangible or intangible. In particular, the Collaborators make no express or implied warranty as to the merchantability or fitness for a particular purpose of any research, Invention, or product, whether tangible or intangible. Likewise, the Collaborators make no express or implied warranty as to any Cooperative Work, Subject Invention, Subject Data, or other product resulting from the Cooperative Work.

4.3.2 The use and dissemination of Information and materials exchanged under this Agreement will be in accordance with all U.S. laws and regulations, including those pertaining to national security and export control. Nothing in this Agreement shall be construed as a license to export Information or to permit any disclosure in violation of law, regulation, or Department of Defense policy. The exporting Collaborator is responsible for obtaining any export licenses that may be required by U.S. Federal law.

Article 5. FUNDING

5.1 Payment Schedule

Consistent with the information in Appendix A, **NAVAIR ORLANDO TSD** will provide source code/databases developed or enhanced by **NAVAIR ORLANDO TSD** to the **MURI / OPUS** lab at no cost to **UCF**. If, at a later date during the period of this CRADA, there is a mutually determined need for additional expertise from **NAVAIR ORLANDO TSD**, then the following will apply. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees to pay **NAVAIR ORLANDO TSD** on a cost basis, as mutually determined to be required, during the period of this CRADA, as specified in Appendix A.

Checks will be payable to:

NAVAIR ORLANDO TSD, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments

Each check and its cover correspondence shall refer to Navy CRADA number "NCRADA-**NAVAIR ORLANDO TSD-02-037**."

Checks will be mailed to:

NAVAIR ORLANDO TSD, 12350 Research Parkway (AIR 4.9T Nimmo), Orlando, Florida 32826-3275, or as otherwise directed by NAVAIR ORLANDO TSD, for specific payments.

5.2 Insufficient and Excess Funds

NAVAIR ORLANDO TSD may discontinue performance under this Agreement if the funds provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** for performance by **NAVAIR ORLANDO TSD** are insufficient or are not provided as specified in Article 5.1. In the event **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to tender the Government the required payment within fifteen (15) days after its respective due date, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall be in default under this Agreement for failure to make payments. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on**

behalf of the MURI / OPUS Laboratory is in default for this reason, NAVAIR ORLANDO TSD shall notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory.** If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** does not cure the default within fifteen (15) days of mailing date of notice, **NAVAIR ORLANDO TSD** may proceed to terminate the Agreement in accordance with Article 11.2.2, may cancel any option for an Exclusive License to a Subject Invention, and may terminate any Exclusive License granted pursuant to this Agreement.

Funds that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** paid under Article 5.1 and that **NAVAIR ORLANDO TSD** has not obligated or expended at the time of completion, expiration, or termination of this Agreement shall be returned to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** after **NAVAIR ORLANDO TSD's** submission of a final fiscal report to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory.**

5.3 No New Commitments

NAVAIR ORLANDO TSD shall make no new commitments concerning this Agreement after receipt of a written termination notice from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in accordance with Article 11.2 and shall, to the extent practicable, cancel all outstanding commitments by the termination date. Should such cancellation result in any costs incurred by **NAVAIR ORLANDO TSD, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that such costs shall be chargeable against any funding that it provided to **NAVAIR ORLANDO TSD.**

5.4 Accounting Records

NAVAIR ORLANDO TSD shall maintain current accounts, records, and other evidence supporting all its expenditures against funding provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement and shall retain such records for at least twelve (12) months after the completion, expiration, or termination of this Agreement. **NAVAIR ORLANDO TSD** shall provide **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** a financial report within four (4) months after completion, expiration, or termination of this Agreement.

Article 6. REPORTS AND PUBLICATIONS

6.1 Final Report

The Collaborators shall submit to each other a final report within four (4) months of the completion, termination, or expiration of this Agreement that includes the results obtained and a list of all Subject Inventions Made.

6.2 Agreement to Confer Prior to Publication or Public Disclosure

The Collaborators agree to confer and consult prior to any publication or public disclosure of Subject Data to ensure that no Proprietary Information, Restricted Access Information, Government Classified Information, CUI, or MCT Information is released and that patent rights are not compromised. Prior to any such publication or public disclosure of Subject Data, each Collaborator shall be offered a period not less than fifteen (15) days and not to exceed thirty (30) days, unless otherwise mutually agreed in writing by the Collaborators, to review any proposed abstract, publication, presentation, or other document for public disclosure that contains Subject Data. For the purposes of this Article, the term "disclosure" shall include, but not be limited to, submission of any manuscript for peer review prior to publication. It is the responsibility of the Collaborator intending to make public disclosure of Subject Data to notify the other Collaborator of such intent.

If a Collaborator objects to a proposed public disclosure, that Collaborator must so notify the other Collaborator within thirty (30) days of the date of notice of intent to disclose publicly. If no objection is received by the Collaborator intending to make public disclosure, concurrence is assumed. If a Collaborator objects on the grounds that patent rights may be compromised, a Patent Application must be filed by the responsible Collaborator within ninety (90) days of the date of notification of intent to make public disclosure, or by another date mutually agreed to by the Collaborators. If a Collaborator objects to the release of Information on the grounds that the Information is Proprietary Information, Restricted Access Information, or Information whose dissemination is restricted by U.S. security laws or regulations, the disclosure shall be postponed until the Information no longer meets the definitions of Proprietary Information, Restricted Access Information, or is no longer covered by U.S. security laws or regulations.

6.3 Classified Information

Any presentation that includes Subject Data that are Classified Information or otherwise restricted Data must have prior review and approval by **NAVAIR ORLANDO TSD** pursuant to the pertinent security laws, regulations, and directives.

Article 7. INTELLECTUAL PROPERTY

7.1 Data

7.1.1 General Provisions Applying to All Data

7.1.1.1 Ownership

Each Collaborator shall have title to all Data generated by that Collaborator.

7.1.1.2 No Implied License

Unless otherwise specifically provided, the Collaborators agree that the exchange of Data of any kind does not confer a license to any Invention claimed in any

patent or Patent Application or to the subject matter of any copyright, trademark/service mark, or other form of Intellectual Property protection.

7.1.1.3 Marking of Data

7.1.1.3.1 Data Provided With Less Than Unlimited Rights

Each Collaborator shall mark all Data that it provides with less than Unlimited Rights with a marking that clearly identifies the limited rights.

7.1.1.3.2 Data That are Proprietary Information or Restricted Access Information

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall place a proper proprietary marking on each medium used for recording Data that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** delivers to **NAVAIR ORLANDO TSD** under this Agreement that **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** asserts is Proprietary Information. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall request in writing if it wishes Subject Data generated by **NAVAIR ORLANDO TSD** to be marked as Restricted Access Information. The Collaborators together shall confer to determine if such marking is appropriate, with reference to the Definitions of Article 1. If the Collaborators mutually agree to the marking then:

(a) For Non-Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory – NAVAIR ORLANDO TSD** MAY USE ONLY FOR PURPOSE OF CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(b) For Subject Data that are Proprietary Information, the marking shall read:

"PROPRIETARY INFORMATION OF **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** – GOVERNMENT HAS GOVERNMENT PURPOSE RIGHTS UNDER CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**";

(c) For Data that are Restricted Access Information, the marking shall read:

"RESTRICTED ACCESS INFORMATION – PROTECT IN ACCORDANCE WITH CRADA NUMBER **NCRADA-NAVAIR ORLANDO TSD-02-037**" UNTIL **SEPTEMBER 2005**".

7.1.1.3.3 Data That are Subject to 35 USC 205

NAVAIR ORLANDO TSD shall mark Data it provides under this Agreement that disclose one or more Inventions in which the Government owns or may own a right, title or interest, and that are subject to confidentiality under 35 USC 205. Such Data shall be marked:

"**NAVAIR ORLANDO TSD** DATA PROTECTED FROM RELEASE OR DISCLOSURE UNDER 35 USC 205."

7.1.1.3.4 Data That are Classified Information, CUI, MCT, or Otherwise Restricted

Each Collaborator shall mark all Data that are Classified Information, CUI, MCT, or otherwise restricted by U.S. security or export control laws or regulations that it provides under this Agreement.

7.1.1.4 Protection of Data

Except for the rights granted in Article 7.1.2.2, Data shall be protected in accordance with the proper markings of its owner and as provided by, at a minimum, the requirements of 15 USC 3710a. Proprietary Information will be protected only if it is properly marked as such. Information provided in intangible form that is Proprietary Information must be designated Proprietary Information at the time it is delivered, followed within fifteen (15) days by a writing summarizing the exact Information to be protected. The Collaborator receiving Information in an intangible form that is designated as Proprietary Information shall be responsible for protecting the Information as Proprietary Information during the fifteen (15) day notification period. After the fifteen (15) day period, if no written summary has been received, the receiving Collaborator need not continue to protect the Information received in intangible form.

Restricted Access Information shall be protected from public dissemination for up to five (5) years, as mutually agreed.

Classified Information, CUI, MCT, or otherwise restricted Information shall be protected in accordance with the security laws of the United States.

7.1.1.5 Release of Data Under the Freedom of Information Act

Data in the possession of **NAVAIR ORLANDO TSD** that are not marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or Restricted Access Information must be released by **NAVAIR ORLANDO TSD** where such release is required pursuant to a request under the Freedom of Information Act (FOIA) (5 USC 552). **NAVAIR ORLANDO TSD** shall protect Data that are properly marked CUI, Proprietary Information of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI**

/ **OPUS Laboratory** or Restricted Access Information from release under the FOIA for as long as the marked Data meet the definition of CUI, Proprietary Information or Restricted Access Information. Prior to release of any such Data, **NAVAIR ORLANDO TSD** shall promptly notify **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** of any request for Data of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** regardless of whether the requested Data are marked Proprietary Information or as required under the public records law.

7.1.2 Subject Data

7.1.2.1 Delivery of Requested Subject Data

Each Collaborator shall have the right to review and receive delivery of all Subject Data generated by the other Collaborator. Requested Subject Data shall be delivered to the requesting Collaborator within fifteen (15) days of the request.

7.1.2.2 Rights in Subject Data

Except as represented in Article 4.3.2, the Collaborators shall have Unlimited Rights in all Subject Data that are not Proprietary Information or Restricted Access Information. Notwithstanding 15 USC 3710a, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** grants Government Purpose Rights in any Subject Data furnished by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to **NAVAIR ORLANDO TSD** under this Agreement that are properly marked as Proprietary Information. The Government has Government Purpose Rights in Subject Data that are Restricted Access Information.

7.1.3 Rights in Non-Subject Data

The Collaborators shall have Unlimited Rights in any Non-Subject Data delivered under this Agreement that are not Proprietary Information.

NAVAIR ORLANDO TSD has a limited right to use, reproduce, and disclose only to Government employees for use in support of the Cooperative Work any Non-Subject Data that are properly marked as Proprietary Information and are provided by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** under this Agreement. Such Proprietary Information can be used only for the purpose of performing the Cooperative Work unless consent to other use or disclosure is obtained from **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in writing.

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall have a limited right to use, reproduce, or disclose Non-Subject Data that may describe one or more Inventions in which the Government owns or may own a right, title or interest, if such Non-Subject Data are provided by **NAVAIR ORLANDO TSD** under this Agreement. In accordance with 35 USC 205, such Non-Subject Data are to be held in confidence. Such Non-Subject Data shall be properly marked by **NAVAIR ORLANDO TSD** and the limited rights of **The UCF Office of Research, UCF on behalf of its**

Board of Trustees on behalf of the MURI / OPUS Laboratory shall be defined by a separate non-disclosure agreement.

7.2 Copyrights

7.2.1 Copyright by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory may copyright works of authorship prepared pursuant to this Agreement if eligible for copyright protection under Title 17 USC.

7.2.2 Copyright License to the Government

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory grants to the Government a nonexclusive, irrevocable, paid-up license in copyrighted works of authorship, including software (17 USC 106) prepared pursuant to this Agreement for any purpose, consistent with the rights in Data described in Article 7.1.

7.2.3 Copyright Statement

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory shall include the following statement on any text, drawing, mask work or other work of authorship, that may be copyrighted under 17 USC, that is created in the performance of this Agreement:

"The U.S. Government has a copyright license in this work pursuant to a Cooperative Research and Development Agreement with **NAVAIR ORLANDO TSD**."

7.3 Trademarks and Service Marks

7.3.1 Ownership of Trademarks and Service Marks

The Collaborator first establishing a trademark or service mark for goods or services with which the mark is used shall be considered the owner of the mark.

7.3.2 Obligation of Employees to Report Trademarks and Service Marks

Employees of both Collaborators shall report the adoption of a trademark or service mark associated with the Cooperative Work to their employer within thirty (30) days of the first use of the mark. Use includes internal use of any product or service of the Cooperative Work.

7.3.3 Obligation of Collaborators to Notify Each Other

Each Collaborator shall notify the other Collaborator within thirty (30) days of their employee's report of the first use of a trademark or service mark.

7.3.4 Responsibility for Filing an Application for Trademark or Service Mark

The Collaborator owning a trademark or service mark shall establish the use of the mark in intra- and interstate commerce and shall be responsible for filing all applications for trademark or service mark registration as appropriate.

7.3.5 License to Use Trademark or Service Mark

The Collaborator owning the trademark or service mark as defined in Article 7.3.1, shall grant a paid-up, irrevocable, nonexclusive license to the other Collaborator for use of the trademark or service mark on the goods or services for which the mark is intended to be used.

7.4 Subject Inventions

7.4.1 Obligation to Report Subject Inventions

7.4.1.1 Collaborators' Instructions to Employees

Each Collaborator shall instruct its employees to submit an Invention Disclosure to that Collaborator for all innovations, solutions to technical problems, or unique increases to the general body of knowledge resulting from the Cooperative Work. For the purposes of this Article, these innovations, solutions, and increases to knowledge shall be deemed Inventions.

7.4.1.2 Timely Invention Disclosure by Inventors

Within ninety (90) days of Making an Invention resulting from the Cooperative Work, unless a shorter time period is required by circumstances, the inventor(s) shall submit an Invention Disclosure to their employer.

In the case of an Invention Made jointly by inventors from both Collaborators, the inventors shall submit an Invention Disclosure with their respective employer.

7.4.1.3 Obligation to Provide Invention Disclosures to the Other Collaborator

Each Collaborator shall provide the other Collaborator with a copy of each Invention Disclosure reporting a Subject Invention within sixty (60) days of receiving the Invention Disclosure from its inventor(s).

7.4.2 Determination of Subject Inventions

The Collaborators shall review each Invention Disclosure resulting from the Cooperative Work and shall confer and consult to determine whether an Invention Disclosure represents a Subject Invention.

7.4.3 Title to and Ownership of Subject Inventions

Each Collaborator shall be entitled to own the Subject Inventions of its employees. Each Collaborator shall cooperate with the other Collaborator to obtain inventor signatures on Patent Applications, assignments or other documents required to secure Intellectual Property protection. For any Invention Made jointly by employees of the Collaborators, each Collaborator shall have ownership of the Subject Invention in the form of an undivided interest.

7.4.4 Filing of Patent Applications

7.4.4.1 Filing of Patent Applications on Solely Made Inventions

Each Collaborator has primary responsibility for filing Patent Applications on the Subject Inventions of its employee(s).

Notwithstanding such primary responsibility, by mutual agreement, the Collaborators may identify which Collaborator shall file a Patent Application on any Subject Invention.

7.4.4.2 Filing of Patent Applications on Jointly Made Inventions

In the case of an Invention jointly Made by employees of both Collaborators, the Collaborators shall confer and agree as to which Collaborator will file any Patent Application. Officers of the non-filing Collaborator shall cooperate with the filing Collaborator to obtain signatures on documents that are needed to file a Patent Application.

7.4.4.3 Preserving Intellectual Property Rights

The Collaborator responsible for filing of a Patent Application on any Subject Invention shall file such Patent Application at least sixty (60) days prior to any bar date or one year from the date the Invention Disclosure was received, whichever comes first. If no Patent Application is filed within the specified time period, the other Collaborator may assume control of filing the Patent Application and take title to the Subject Invention on ten (10) days written notification. The Collaborator that relinquished the responsibility to file shall retain a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Subject Invention practiced throughout the world by or on its behalf.

7.4.4.4. Filing Deadlines

The Collaborator responsible for filing any Patent Application for a Subject Invention shall notify the other Collaborator of all filing deadlines for prosecution of any Patent Application and maintenance of any patents on the Subject Invention. Notwithstanding the primary responsibility defined in Article 7.4.4.1, sixty (60) days prior to any filing deadline, the Collaborators shall confer to determine if the filing Collaborator intends to respond to the filing deadline. The non-filing Collaborator will be permitted to take action if the filing Collaborator declines.

7.4.4.5 Copies and Inspection

7.4.4.5.1 Copies of Prosecution Papers

Each Collaborator filing a Patent Application on a Subject Invention shall provide the other Collaborator with a copy of any communication relating

to prosecution of said Patent Application within thirty (30) days of receipt of such communication.

7.4.4.5.2 Access to Patent Application File and Right to Make Copies

Upon written request, the filing Collaborator shall give the other Collaborator an Associate Power of Attorney, with authorization to access the Patent Application, make copies, and, in the event the filing Collaborator fails or declines to take action, do all that is necessary to secure Intellectual Property protection for the Subject Invention.

7.4.4.6 Rights of Inventors if the Collaborators Decline to File a Patent Application

In the event both Collaborators decline to file a Patent Application on a Subject Invention, the Government will renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

In the event both Collaborators decline to file a Patent Application on a Subject Invention, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** may, at its sole discretion, renounce its entitlement and leave its rights to the inventor(s) who may retain ownership of the Invention, subject to the retention by each Collaborator of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have the Invention practiced throughout the world by or on its behalf.

7.4.5 Nonexclusive License to Subject Inventions

7.4.5.1 Nonexclusive License Grant

Each Collaborator grants to the other Collaborator a nonexclusive, irrevocable, paid-up license to practice a Subject Invention Made by employees of the granting Collaborator or have the Subject Invention practiced throughout the world by or on behalf of the other Collaborator. No nonexclusive license granted under this Agreement shall permit licensee to grant sublicenses.

7.4.5.2 Confirmatory Nonexclusive License Agreement

Each Collaborator has the obligation to provide a Confirmatory License Agreement (Appendix B) to the other Collaborator for each nonexclusive license within ninety (90) days of the date of filing.

7.4.6 Option for Exclusive License to Subject Inventions

NAVAIR ORLANDO TSD gives **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** the option of acquiring an Exclusive License for the field of use **(virtual environments for operator/small arms training)** in the Government's rights in any Subject Invention Made in whole or in part by a **NAVAIR ORLANDO TSD** employee. The license shall be for reasonable consideration. In

order to exercise this option, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must notify **NAVAIR ORLANDO TSD** in writing within one hundred and eighty (180) days of the filing of a Patent Application. Unless another time period is mutually agreed upon between the Collaborators, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** must execute an Exclusive License to the Subject Invention within one hundred and eighty (180) days of election to exercise the option, or the Invention shall be made available for licensing by the public in accordance with 37 CFR Part 404.

Any Exclusive License granted by the Government in a Subject Invention is subject to the statutorily required reservation by the Government of a nonexclusive, irrevocable, paid-up license to practice the Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the Government (15 USC 3710a).

7.4.7 Limitation on Assignment of Licenses Granted Under This Agreement

No license granted under this Agreement shall be assigned, licensed or otherwise disposed of except to the successor in interest of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which such license pertains.

7.4.8 Termination of License Granted and Cancellation of Exclusive License Option to Subject Inventions

7.4.8.1 Exclusive Licenses and Exclusive License Option

NAVAIR ORLANDO TSD may terminate any Exclusive License or cancel any option for an Exclusive License to a Subject Invention granted under this Agreement in the event that:

(a) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** is in default for failure to make payment as agreed in Article 5; or

(b) The Agreement is terminated unilaterally by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**; or

(c) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** fails to perform according to the Statement of Work (Appendix A); or

(d) **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** becomes a foreign owned, controlled, or influenced (FOCI) organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.4.8.2 Nonexclusive Licenses

NAVAIR ORLANDO TSD shall terminate any nonexclusive license to a Subject Invention granted under this Agreement if **The UCF Office of Research,**

UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory becomes a FOCI organization that does not qualify under the requirements of Executive Order 12591, Section 4(a).

7.5 Non-Subject Inventions

7.5.1 Ownership of Non-Subject Inventions

Each Collaborator owns its Non-Subject Inventions.

7.5.2 Rights Under Other Agreements

Nothing in this Agreement is intended to change the rights in Intellectual Property acquired by the Collaborators in any other contract or agreement between the **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** and the Government.

7.5.3 No License to Non-Subject Inventions

This Agreement does not grant any Collaborator a license, express or implied, to any Non-Subject Invention.

7.6 Research License

Each Collaborator shall allow the other Collaborator to practice any of its Non-Subject Inventions for the purpose of performing the Cooperative Work.

No license, express or implied, for commercial application(s) is granted to either Collaborator in Non-Subject Inventions by performing the Cooperative Work.

For commercial application(s) of Non-Subject Inventions, a license must be obtained from the owner.

Article 8. TANGIBLE PROPERTY

8.1 Title to Preexisting Tangible Property

Each Collaborator shall retain title to all Tangible Property to which it had title prior to the Effective Date of this Agreement.

8.2 Tangible Property Purchased by Collaborators to Perform the Cooperative Work

Each Collaborator shall retain title to all Tangible Property that it purchases during the period of this Agreement. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** cannot take title to any Government Tangible Property under this Agreement. Collaborator consumables to be used in the Cooperative Work of this Agreement are the property of the purchasing Collaborator until consumed.

8.3 Title to Developed Tangible Property

All Tangible Property developed under this Agreement with all components purchased by one Collaborator shall be the property of that Collaborator. Tangible Property having any component purchased by **NAVAIR ORLANDO TSD** shall be the property of the Government, unless such Tangible Property can reasonably be separated without damage to the other individual components. After this Agreement is completed, expired, or terminated, if separation of components can be made without damage, the Collaborators may, by mutual agreement, separate the Tangible Property into its components and the separated components shall remain the property of the Collaborator that purchased them.

8.4 Tangible Property Operational and Disposition Costs

During the period of and upon completion, expiration, or termination of this Agreement, each Collaborator shall be responsible for all costs of maintenance, removal, storage, repair, disposal, and shipping of all Tangible Property to which it has title.

8.5 Disposal of Tangible Property

Disposal of Tangible Property shall be in accordance with applicable U.S. Federal, State, and local property disposal laws, environmental laws, and regulations.

Article 9. LIABILITY

9.1 Extent of Government Liability

The Government shall be solely liable for the negligent or wrongful acts of its officers and employees to the extent provided for in the Federal Tort Claims Act (28 USC 2671 et. seq.) and in other applicable laws and regulations of the United States that specifically waive sovereign immunity. Nothing in this Agreement shall be construed as a waiver of the sovereign immunity of the United States.

9.2 Extent of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** Liability

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory is solely responsible for its actions and the actions of those acting for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** in the performance of this Agreement and for any damages that may arise from any suit, action, or claim, and for any costs from or incidental to any suit, action, or claim, including but not limited to settlement and defense costs. Further, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agrees that in any suit, action or claim brought by anyone not a party to this Agreement based on actions of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory, The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not pursue any actions to enter the Government as a party in such suit, action or claim unless the Government has some liability under the Federal Tort Claims Act.

9.3 *Force Majeure*

No Collaborator shall be liable for the consequences of any *force majeure* that (1) is beyond its reasonable control; (2) is not caused by the fault or negligence of such Collaborator; (3) causes such Collaborator to be unable to perform its obligations under this Agreement; and (4) cannot be overcome by the exercise of due diligence. In the event of the occurrence of a *force majeure*, the Collaborator unable to perform shall promptly notify the other Collaborator. The Collaborators shall suspend performance only for such period of time as is necessary to overcome the result(s) of the *force majeure* and shall use their best efforts to resume performance as quickly as possible.

Article 10. GENERAL PROVISIONS

10.1 Characteristics of the Agreement

10.1.1 Entire Agreement

This Agreement constitutes the entire agreement between the Collaborators concerning the Cooperative Work and supersedes any prior understanding or written or oral agreement relative to the Cooperative Work.

10.1.2 Severability

The illegality or invalidity of any Article of this Agreement shall not impair, affect, or invalidate any other Article of this Agreement.

10.1.3 Interpretation of Headings

Headings of the Articles of this Agreement are for convenience of reference only and do not form a part of this Agreement and shall in no way affect the interpretation thereof.

10.2 Agreements Between Collaborators

10.2.1 Governing Laws

United States Federal Laws shall govern this Agreement for all purposes.

10.2.2 Independent Parties/Entities

The relationship of the Collaborators to this Agreement is that of independent parties and not as agents of each other, partners, or participants in a joint venture. Each Collaborator shall maintain sole and exclusive control over its personnel and operations.

10.2.3 Assignment/Subcontracting

10.2.3.1 Neither Collaborator may allow third parties to perform any part of the Cooperative Work under this Agreement without express written consent of the other

Collaborator. If consent is obtained, the Collaborator requesting such consent shall remain fully responsible for the portion of the Cooperative Work to be accomplished under a third-party agreement, and the third party is not a Collaborator of this Agreement. Any third-party agreement to perform a portion of the Cooperative Work shall contain terms consistent with this Agreement.

10.2.3.2 This Agreement shall not be assigned or otherwise transferred by either Collaborator without the prior written consent of the other Collaborator, except to the successor of that part of **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**'s business to which this Agreement pertains.

10.2.3.3 If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee is a U.S. company, and becomes, during the term of this Agreement or thereafter, directly or indirectly owned, controlled, or influenced by a foreign company or government (FOCI), then **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** or its successor or assignee shall promptly notify **NAVAIR ORLANDO TSD** to that effect.

10.2.4 Disputes

10.2.4.1 Settlement and Resolution

NAVAIR ORLANDO TSD and **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** agree to use reasonable efforts to reach a fair settlement of any dispute. If such efforts are unsuccessful, remaining issues in dispute will be referred to the signatories or their successors for resolution. If a dispute continues, the remaining issues may be submitted to the Chief of Naval Research (CNR), or the CNR designee, for resolution. This Agreement does not prevent any Collaborator from pursuing disputes in a U.S. Federal court of competent jurisdiction. No Collaborator will pursue litigation in a U.S. Federal court until after the CNR, or the CNR designee, decides the dispute, or until sixty (60) days after the dispute was first submitted to the CNR, or the CNR designee, whichever comes first.

10.2.4.2 Continuation of Cooperative Work

If payments or installment payments are to be made as stated under Article 5, **NAVAIR ORLANDO TSD** will not start or continue cooperative work until payments or installment payments are received.

10.2.5 Waivers

None of the provisions of this Agreement shall be considered waived by either Collaborator unless such waiver is given in writing to the other Collaborator, signed by the executing official of this Agreement or the official's successor having the authority to bind the Collaborator making the waiver. The failure of either Collaborator to insist upon strict performance of any of the terms and conditions herein, or failure or delay to exercise any rights provided herein or by law shall not be deemed a waiver of any right of either Collaborator under this Agreement.

10.2.6 Use of Name or Endorsements

Except as provided for in Article 7.2.3, **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not use the name of **NAVAIR ORLANDO TSD** or any other Government entity on any product or service that is directly or indirectly related to either this Agreement or any patent license or assignment associated with this Agreement without the prior approval of **NAVAIR ORLANDO TSD**. By entering into this Agreement, **NAVAIR ORLANDO TSD** does not directly or indirectly endorse any product or service provided, or to be provided, by **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, its successors, assignees, or licensees. **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** shall not in any way imply that the Department of the Navy endorses any such product or service.

10.3 Environment, Safety, and Health

Each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste in its custody during the course of this Agreement. At the conclusion of this Agreement, each Collaborator shall be responsible for the handling, control, and disposition of any and all hazardous substances or waste still in its possession. Each Collaborator shall obtain at its own expense all necessary permits and licenses as required by U.S. Federal, State, and local law and shall conduct such handling, control, and disposition in a lawful and environmentally responsible manner. Each Collaborator is responsible for all required environmental, safety, and health compliance, notice, and monitoring related to its facility in accordance with U.S. Federal, State, and local law and regulations. Collaborators shall abide by the environmental, safety, and health directives of the host facility in which the Cooperative Work is being performed, and any U.S. Federal, State, or local laws and regulations pertaining to environment, safety, and health that are applicable to the host facility.

10.4 U.S. Competitiveness

The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory agrees that any product, process, or service using Intellectual Property arising from the performance of this Agreement shall be manufactured substantially in the United States.

10.5 Public Release of This Agreement

This Agreement, without funding information (Article 5) and Appendices, may be released to the public.

Article 11. MODIFICATIONS AND NOTICES

11.1 Amendments

If a Collaborator wishes to modify this Agreement, the Collaborators shall confer in good faith to determine the desirability of such modification. Such modification shall not be effective until a written amendment is signed by both executing officials of this Agreement or their successors.

11.2 Termination

11.2.1 Termination by Mutual Consent

The Collaborators may elect to terminate this Agreement at any time by mutual consent. Such termination shall not be effective until a written termination agreement is signed by both executing officials of this Agreement or their successors.

11.2.2 Unilateral Termination

A Collaborator may unilaterally terminate this entire Agreement at any time by giving the other Collaborator written notice signed by the executing official of this Agreement or his/her successor, not less than thirty (30) days prior to the desired termination date. If **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** unilaterally terminates this Agreement, any option for an Exclusive License to a Subject Invention and any Exclusive License to a Subject Invention granted by or pursuant to this Agreement shall simultaneously be terminated.

11.3 Notices

All notices pertaining to or required by Articles of this Agreement, except those pertaining solely to the prosecution of any patent, trademark, or service mark, shall be in writing and shall be signed by an authorized representative of the Technology Transfer Office for **NAVAIR ORLANDO TSD** or the preferred contact for **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**, and all such notices shall be delivered by hand, sent by courier with proper registration, or sent by certified mail, return receipt requested, with postage prepaid, addressed as follows:

If to **NAVAIR ORLANDO TSD**:

Ms. Patricia R. Nimmo
Technology Transfer Project Manager
NAVAIR ORLANDO TSD (AIR 49T)
12350 Research Parkway
Orlando, FL 32826-3275

If to **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

Ms. Kim Smith
Senior Contract Specialist
Office of Research
University of Central Florida
Orlando Tech Center
12443 Research Parkway, Suite 207
Orlando, Fl 32826-3252

A Collaborator shall notify the other Collaborator of a change of address in the manner set forth above.

Notices pertaining solely to the prosecution of any patent, trademark, or service mark related to this Agreement shall be in writing and shall be signed by and sent to the Collaborator's legal counsel for Intellectual Property. Legal counsel for Intellectual Property for each Collaborator shall send a copy of any such notice to the Technology Transfer Office for he **NAVAIR ORLANDO TSD**. If either Collaborator fails to identify such counsel upon request, then such notices shall be sent to the points of contact specified above.

Article 12. SURVIVING PROVISIONS

The Articles covering Definitions, Representations and Warranties, Funding, Reports and Publications, Intellectual Property, Tangible Property, Liability, General Provisions, Modifications and Notices, and Surviving Provisions shall survive the completion, termination, or expiration of this Agreement.

Article 13. DURATION

This Agreement expires three years after its Effective Date, unless otherwise extended in writing according to the provisions of Article 11.

Article 14. SIGNATURES

For **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory**:

I, the undersigned, am duly authorized to bind **The UCF Office of Research, UCF on behalf of its Board of Trustees on behalf of the MURI / OPUS Laboratory** to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002

By: M.J. Soileau

Title: Vice President, Office of Research

For the Department of the Navy:

I, the undersigned, by 15 USC 3710a and Navy regulations, am duly authorized to bind the U.S. Navy to this Agreement and do so by affixing my signature hereto.

Entered into this 30 day of Sep 2002.

By:

R. A. MOHLER
Title: Captain, U.S. Navy
Commanding Officer

Navy Organization: NAVAIR ORLANDO, TRAINING SYSTEMS DIVISION

This Page Intentionally Left Blank.

APPENDIX A

STATEMENT OF WORK
BETWEEN
NAVAIR ORLANDO
TRAINING SYSTEMS DIVISION
AND
THE UNIVERSITY OF CENTRAL FLORIDA (UCF) OFFICE OF RESEARCH, UNIVERSITY OF CENTRAL FLORIDA (UCF)
ON BEHALF OF
ITS BOARD OF TRUSTEES
ON BEHALF OF
THE
UCF MULTIDISCIPLINARY RESEARCH PROGRAM OF THE UNIVERSITY RESEARCH INITIATIVE / OPERATOR PERFORMANCE UNDER STRESS (MURI / OPUS) LABORATORY

NAVAIR ORLANDO TSD will:

1. Provide, for use in the UCF MURI/OPUS laboratory, source code/databases developed or enhanced by NAVAIR ORLANDO TSD as part of Navy sponsored research and development, e.g., Host Software, Image Generator Software, Weapon Tracking Software, Networking/Communications Software, Data Collection/Analysis Software, Scenario Builder Software, Rural and Urban Databases

2. Provide expertise, on a cost basis and as mutually determined to be required, on utilization of government provided software and databases for UCF MURI/OPUS Laboratory research

3. Provide, on a cost basis and as mutually determined to be required, expertise for further hardware/software/database development, testing, and training applications

4. Participate, on a cost basis, in mutually selected specific research projects conducted under the auspices of the UCF MURI/OPUS Laboratory

5. Provide access to NAVAIR ORLANDO TSD customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

6. Identify potential military training applications for shared research findings

UCF MURI/OPUS Laboratory will:

1. Provide laboratory space, equipment and facilities required to house and utilize government provided software and databases

2. Purchase any necessary commercial software and licenses required to create, build, and run government provided source code and databases

3. Provide funding, on a cost basis and as mutually determined to be required, for NAVAIR ORLANDO TSD expertise and assistance

4. Assume responsibility for UCF MURI/OPUS Laboratory system validation and data collection.

5. Conduct an annual meeting highlighting significant findings and provide copies of all publications resulting from research in the UCF MURI/OPUS Laboratory

6. Provide NAVAIR ORLANDO TSD with access to MURI/OPUS Laboratory enhancements to subject data

7. Assume responsibility for MURI/OPUS Laboratory maintenance and repair costs related to activities of this CRADA

NAVAIR ORLANDO TSD and UCF MURI/OPUS Laboratory will:

1. Identify mutually beneficial projects suitable for collaborative research

2. NAVAIR ORLANDO TSD and the MURI/OPUS laboratory will, by mutual consent, coordinate activities to facilitate laboratory demonstrations for a wide audience of potential researchers and users, and will share mutual access to customers and organizational partners with interest and expertise in Human Factors for homeland security and related technologies

3. Mutually develop processes separate from this agreement to allow exchange of funds, should that be mutually determined to be required during the period of this CRADA

NOTE 1: The parties agree that Article 9.2 Extent of Liability, will be subject to and interpreted in accordance with Florida Statutes, Section 768.28.

NOTE 2: FOREIGN ACCESS TO TECHNOLOGY

All work described in this SOW shall be subject to the following with regard to foreign access to technology.

The Collaborators agree that research findings and technology developments under this CRADA may constitute a significant enhancement to the national defense. Accordingly, access to important technology developments under this CRADA by foreign firms or institutions must be carefully controlled. These controls are in addition to, and are not intended to change or supercede, the provisions of the International Traffic in Arms Regulation (22 CFR pt. 121 et seq.) the DoD Industrial Security Regulation (DoD 5220.22-R) and the Department of Commerce Export Regulation (15 CFR pt.770 et seq.)

The non-Navy Collaborator shall provide timely notice to the government's designated Principal Investigator of any proposed transfer of research findings and technology developments under this CRADA to foreign firms or institutions. If the government determines that the transfer may have adverse consequences to the national security interests of the United States, the non-Navy Collaborator and the government shall jointly endeavor to find alternatives to the proposed transfer which obviate or mitigate potential adverse consequences of the transfer but which provide substantially equivalent benefits to the non-Navy Collaborator.

APPENDIX B

CONFIRMATORY LICENSE AGREEMENT	*1. APPLICATION FOR (Title of Invention)*
2. INVENTOR(S) AND AFFILIATION	
3. PATENT APPLICATION SERIAL NO.	*4. PATENT APPLICATION FILING DATE*
5. NAVY ACTIVITY (Name, address, point of contact)	*6. NON-NAVY ACTIVITY (Name, address, point of contact)*
7. CRADA AGREEMENT NO.	*8. DATE OF THIS AGREEMENT*

9. The Invention identified above is a "Subject Invention" under Article 7 Intellectual Property included with the CRADA identified in Box 7 between the Department of the Navy and Non-Navy Activity identified in Box 6.

This document is confirmatory of the nonexclusive, irrevocable, paid-up license to practice the identified Subject Invention or have that Subject Invention practiced throughout the world by or on behalf of the receiving party, and of all other rights acquired by the receiving party by the referenced clause.

This license is granted to

_____ *the Government*

(Select one)

_____ *Non-Navy Activity identified in Box 6*

under this CRADA in the identified Invention, Patent Application and any resulting patent.

The licensee is hereby granted an irrevocable power to inspect and make copies of the above-identified Patent Application.

_____ *ACTIVITY NAME OF LICENSOR*

_____ *SIGNATURE*

_____ *NAME (Typed or Printed)*

_____ *TITLE*

_____ *BUSINESS TELEPHONE*

This Page Intentionally Left Blank.

MITIGATING STRESS, WORKLOAD, AND FATIGUE ON THE ELECTRONIC BATTLEFIELD

The MURI-OPUS Program

Project Summary

Our military forces are now facing new and often unanticipated challenges Many of the tasks they are now being asked to perform are not those which typically occupy the military mandate. That they do this so well argues for their professionalism and their adaptability but it brings in to play sources of stress and demand for which traditional military training provides little experience or protection. The better the theories and principles that we generate and validate, the better we will be able to help these individuals in their new and evolving missions. It is evident that current and future military operations will be 'cognitive' wars (Scales, 2006). The future battleground will largely be the minds of other individuals and those individuals like our own forces, will certainly be under stress. It is crucial that we know about these effects if we are to understand them fully and exploit them effectively. Our Multidisciplinary University Research Initiative (*MURI*) program on Operator Performance Under Stress (*OPUS*) has contributed substantially to the scientific and operational understanding of stress and performance. This MURI reached completion in December 2006. Here we summarize the outcome of this program and the numerous successes achieved by researchers funded under this MURI.

Minds in Technology, Machines in Thought (MIT^2) Laboratory

At the University of Central Florida, the MIT^2 research group has used a number of research test beds to evaluate the theoretical propositions within the maximal adaptability model of stress and performance as first developed by Hancock and Warm (1989). This has resulted in increased understanding of the variations in spatio-temporal perception under stressful conditions. Specifically, experimental evidence suggests that the two dimensions, which form base axes of the Hancock and Warm (1989) model, each represent distinct perceptual mechanisms. Further, variations in spatial and temporal properties of tasks on stress response depend in part on an individual's personality traits. Specifically, individuals higher in dispositional pessimism reported increased stress symptoms after performing a short signal detection task, but only when that task included spatial uncertainty in stimulus appearance. The effect of these insights is to advance us significantly closer to achieving the goal of quantitative performance

prediction under stress, contingent upon these detailed elaborations of the required performance task and the personnel disposition. This allows us to make recommendations as to selection and training for stressful missions and provide design recommendations for technological support and even task redesign to assure mission success under the highest levels of demand imposed on our forces.

At UCF, we have also established that Fuzzy Signal Detection Theory (FSDT) is a particularly useful quantitative method to understand detection capacities under stress. Experimentally, we have confirmed that FSDT can exploit all of the mathematical advantages of traditional signal detection theory. Our current evidence indicates that observers have difficulty establishing response criteria when judging fuzzy stimuli. This implies that one problem in decision making under stress and uncertainty may be in setting a criterion for responding. Strong evidence for the utility of the FSDT model was established by the use of more complex stimuli that represent real-world military demands of armor identification. Specifically, Receiver Operating Characteristic (ROC) functions were derived using a tank discrimination task using morphed images of M1A1 and T55 tanks. The findings from this experiment confirmed our earlier empirical results using standard laboratory-generated psychophysical stimuli. Further experiments have established that forcing observers to make yes/no decisions reduced their perceptual sensitivity relative to those allowed to make more graded (i.e., rating) judgments. With regard to the latter, evidence indicates that permitting 5 categories of response yields ROC functions similar to those obtained using 100 response categories. This means we can now make a strong effort at quantitative prescription of fuzzy circumstances such as ambiguities in relation to 'rules of engagement.' We have also established the applicability of these detection measures for technology transfer to use in civilian realms such as medical diagnostic applications to monitoring tasks and in law enforcement. These measures add immeasurably to traditional signal detection metrics in the evaluation of performance on detection and threat-evaluation tasks. The criterion issue is particularly relevant to this MURI effort because one of the major sources of stress facing soldiers in combat environments is the uncertainty associated with the threat posed by the objects and events they face during missions. The evidence from the FSDT experiments funded under this MURI indicate that the next step is to identify how specific individuals set these fuzzy criteria and the mechanisms that control their criterion fluctuation as a function of time and changes in the environment. Our research efforts have demonstrated that FSDT is the preferred methodology for resolving these crucial issues.

MIT^2 has also examined the role of individual differences in adaptation to stress in perceptual and cognitive tasks. Evidence to date indicates that the personality traits of individuals influence their stress level and the coping strategies they employ to deal with the stress in a systematic and predictable manner. One variable, *attentional control*, has been shown to be linked to the perception of workload and stress associated with using adaptive automation. In addition, we have found that traits such as *extraversion* and *emotional stability* interact to influence how individuals react to the stress associated with the demands inherent in shooting tasks. MIT^2 researchers have also focused their experimental investigation of the stress of sustained attention, which is extremely relevant to current conflict environments and their associated problems such as the presence of 'improvised explosive devices' (IED's). Our findings have indicated that providing observers with an expectation of control over the kind of vigilance task they experience improves performance significantly however, this is exclusively the case only when the condition they receive matches their preference. Other experimental work has established that the capacity for training with knowledge of results feedback to reduce stress symptoms depends on the *dispositional pessimism* of the observer. These and like findings from other component members of the MURI program demonstrate that it is possible to improve both detection and vigilance on behalf of military personnel and thus improve both safety and mission performance

MIT^2 field-based research has used a variety of analog circumstances and we have taken especial advantage to examine shooting performance at a police training range to which we have gained access. This facility allows us to measure stress in environments appropriate to those occurring for modern soldiers in broken urban settings. Using this facility, we have confirmed the workload-performance association/dissociation patterns predicted in the Hancock and Warm model in the context of a real-world shooting task with professional police officers. Of particular relevance is the fact that these data were inconsistent with the traditional 'inverted U' explanation of performance response under stress, further indicating that this older representation does not reflect nor explain real-world data. Thus MURI has had a profound effect on basic psychological theory as to how stress affects performance capacity. These insights should change how this topic is taught on a world-wide basis. The results of this study also confirmed the importance of considering the subscales of perceived workload measures, and underscore the multidimensionality of the construct itself. Through a Cooperative Research and Development Agreement (*CRADA*) with the Naval Air Warfare Center, Training systems Division, Orlando, the MIT^2 laboratory created, refined,

and exploited a Small Arms Simulation Test Bed (SAST) facility. We conducted numerous experimental investigations of cognitive changes that occur in a stressful target acquisition task. Findings indicated performance on an auditory secondary task, in which the individual must process information regarding the identification and location of other friendly forces, is disrupted by concurrent target acquisition demands. Although shooting accuracy was not significantly impacted by the presence of a secondary task, response time was slower at the highest demand level, indicating a cost for maintaining accuracy under high workload conditions. Further, the performance decrement was observed only on trials in which a target was present. These insights into the information overload problem provide us with strong evidence from which to understand the limits of soldier tasking in high stress conditions. Particularly, it provides important insight into task distribution as new technologies such as UAV support percolate to the squad level.

MIT^2 has established two PC-based test beds for evaluating performance under stress. First, Ghost Recon has been employed for investigation of the effects of time pressure on land navigation tasks. Second, an existing simulation system is being used to test the impact of noise on navigation, monitoring, and time distortion under stress. Results indicated important time distortions under stress, with participants significantly underestimating durations at higher levels of noise. Further studies investigated the effects of cross-modality of warning presentation and retention in dual-task circumstances in a simulated military environment with various task-induced stress levels. These results revealed that when the presentation format and the response format matched (e.g., verbal-verbal), behavioral compliance was greater then when presentation and response format were mismatched (e.g., verbal-pictorial). Furthermore, at the greatest level of working memory demand (eight warnings) the operational task resulted in important performance degradations. Such findings provide important design guidelines for providing emergency warnings to soldiers in the midst of combat.

MIT^2 has also installed a military vehicle simulation for testing hypotheses concerning fatigue and sustained performance. Our findings indicated that how fatigue influences on military driving performance varies as a function of terrain, suggesting adaptive response by drivers over time. That is, drivers are able to adopt different strategies to compensate for performance decrement by narrowing their attention to critical cues in a particular environment. These findings have implications for the development of countermeasures to fatigue. In addition, results of the tests of our driving scenarios supports the use of this task

environment for investigation of workload effects over time. We demonstrated this by showing that high-quality measures of workload could be obtained while driving an equidistant and symmetrical route under varying degrees of workload imposed by task characteristics other than the act of driving itself.

The *VIKING* experimental environment, originally developed at the University of Minnesota under MURI sponsorship, has been used to assess performance on tasks that impose information processing demands equivalent to those faced by soldiers in combat environments. The MIT^2 Laboratory used the more recently developed ARES battery to assess the changes in cognitive performance of soldiers in the field during simulated combat. Findings indicated that many soldiers tolerate such field conditions with minimal impairment of cognitive performance if they know what is expected, but that high demand events are still associated with cognitive performance decrement. In addition, it appears that different forms of cognitive task are not affected equally in the context of combat. Cases were observed that clearly showed a precipitous decline in adaptive function at the extremes of stress, as predicted by the Hancock and Warm (1989) model. The decrement in cognitive function was replicated in a sample of participants in a Survival Evasion, Resistance, and Escape (SERE) exercise, with the more complex cognitive tasks (e.g., logical reasoning) showing the greatest susceptibility to the effects of sustained stress. In addition, performance immediately after the exercise showed a return to normal (pre-exercise) levels, but only for a short period of time. The field work using the ARES has established the utility of the battery of tasks for measuring the effects of stress in operational settings (e.g., field training exercises) on perceptual and cognitive processes.

As well as these on-going empirical efforts, researchers in the MIT^2 laboratory have completed meta-analytic reviews of the effects of temperature, vibration, and noise on human performance. These efforts are geared toward enhancing IMPRINT capacity for prediction of performance under stress. Each of these stressors impairs performance, but in the case of thermal stress, the impact depends on the range of temperature, as well as the nature of the task and the duration of exposure. It is likely that intensity, duration, and task type also moderate the relation of vibration to performance, but there were insufficient studies in the published literature to evaluate these hypotheses. More fine grained analysis of vibration and performance will require programmatic experimental work by future researchers.

As part of our efforts to refine the Hancock and Warm model, in particular the task dimensions, we developed a theoretical framework for the understanding and quantifying the limitations to human information processing. This model is based on different timing mechanisms at different levels in the brain. Incorporated into this framework is the role of emotion, both positive and negative, in information processing mechanisms, and the trait variables that influence this link. We have also elaborated the multiple resource theory 'box model' to improve its utility for both research and practice.

Team Performance Laboratory (TPL)

The Team Performance Laboratory (*TPL*) at the University of Central Florida has investigated adaptability of teams under time pressure and high workload, and is currently using a military-based task (*Flashpoint*) to further investigate the impact of stress on team processes. In particular, derived evidence indicates that the quality of team communication and performance varies as a function of team type (e.g., command and control versus a reconnaissance team). TPL researchers have also found that time pressure impairs communication among team members, breaking the closed-loop communication pattern. Evidence indicates that teamwork behaviors (i.e., communication) and team performance under stress are influenced by task interdependency. These results confirmed that teams should not be treated as equivalent and that team training should incorporate differences in team type and in the degree of task interdependency. In addition to these efforts, TPL researchers have made advances on the development of metrics to assess team performance under conditions of high stress and workload. It is anticipated that these can be transitioned to further research to implement the empirical and theoretical advances of this MURI to developing training regimens for teams, as well as transitions to civilian domains. Toward that end, TPL researchers developed 'best practices' guidelines for team training.

Linköping University (LiU)

MURI researchers at Linköping University (LiU), (formerly associated with Kansas State University) investigated the effects of trust on team decision-making, particularly, the identification and mitigation of cognitive decrements due to remote command and control under realistic simulations of battlefield conditions. These efforts resulted in the completion of a viable research paradigm for the investigation of trust and team problem-solving, and represent one of the myriad examples of cross Laboratory technology transfer and protocol integration. The

LiU group conducted several simulated combat environment experiments on the effects of social interaction on performance. These experiments document a cognitive decrement directly attributable to remote command and control. Military simulations using paintball revealed that when leaders engage in remote command and control there is a decrement in soldiers' trust for the leader. Associated with this loss of trust is compromised performance. Further experimentation using their paradigm indicated that use of vibrotactile stimulation to issue commands remotely can compensate for hesitation to commands from remote commanders. They also established that the use of information and communication technology for emergency management can lead to performance decrements if there is a mismatch in its use between collocated training and distributed emergency management operations. They extended the decision making research described above to the investigation of the effects of cultural differences on decision making efficiency in multicultural command-and-control centers. Their research indicated that culturally-based differences in planning and leadership style can impair decision making efficiency in these contexts.

George Mason University

MURI researchers in the Cognitive Science Laboratory at the George Mason University (formerly associated with the Catholic University of America) have completed a sequence of studies on the impact of automation on performance in decision-making tasks of the kind typically faced by military personnel in command and control settings. They found that as a result of complacency, automation with high (but not perfect) reliability does not always best support performance. They also found that automation unreliability has a greater performance cost for decision automation than for information automation. In addition, they continue to explore the effectiveness of adaptive automation using real-time measurement of workload under conditions of sustained demand and perceived stress. These researchers have also developed methodology for enabling flexible supervisory control that can be delegated to automation, and they have demonstrated how these methods can be applied to UAV control.

University of Cincinnati

Researchers at the University of Cincinnati have used Transcranial Doppler Sonography System (TCDS) to examine the effect of sustained attention on brain blood flow. They conducted experiments demonstrating that the decline in sustained attention performance over time is accompanied by a parallel decline in

cerebral blood flow in the right hemisphere. In addition, they have initial evidence that transcranial oximetry (TCCO) measures of blood flow and oxygenation levels in the brain may also provide a useful physiological measure of sustained attention. In a recent experiment investigating the effect of type of automation support on decision making it was found that performance (both speed and accuracy) in automated tasks improved when changes to information automation were adaptive, and that changing automation type in the middle of a task reduced the cost associated with automated decision aid. Thus, response time under manual control was better after information automation. These results indicated that adaptive changes in automation type during a task can reduce the costs associated with unreliable automation. In particular, changes from recommendation to listing automation formats improves both accuracy and response time, while change from decision recommendation to manual improved response time. Thus, providing some information automation during a session before switching to decision automation can facilitate performance.

UC researchers also compared the effectiveness of a new measure of perceived mental workload, the Multiple Resources Questionnaire (MRQ) to that of the NASA-TLX, which has been established as one of the major instruments for the measurement of perceived workload. The results indicated that the MRQ can be a useful adjunct to the NASA-TLX in measuring the workload of vigilance tasks and that it permits the identification of information processing resources utilized in the performance of vigilance tasks that are not measured by the NASA-TLX. Further, they demonstrated that the level of cerebral blood flow varies directly with task demand, and that these effects were lateralized to the right hemisphere indicating right hemisphere functional control of vigilance performance.

MURI Support of Researchers

In respect of all these achievements, MURI has supported over thirty personnel at the multiple identified universities including a total of nine Institutions in all. Such personnel included five post-doctoral researchers and more than twenty graduate student researchers across all participating members. Several of the upper level students have now graduated with Ph.D's and are continuing their MURI-related and MURI-inspired efforts at a number of locations including academe, industry, and in military contexts. As one representative success of these students, at least two MURI-linked individuals will be teaching at West Point for the US. Army The participating universities in the program were: University of Central Florida, University of Minnesota, Minnesota State University

at Mankato, Kansas State University, Linköping University, University of Cincinnati, Catholic University of America, George Mason University and Montana State University.

Contributions to Scientific Understanding

We have made several fundamental steps in scientific understanding concerning stress effects. We have begun to elucidate the effects of diverse sources of real-world stress on spatio-temporal perception. Using stress sources that will impact further development of the IMPRINT model, we have seen confirmation of our hypotheses on the 'narrowing' of the perceptual world of the stressed operator. These efforts have burgeoned into display design guidelines for such operations and in modifications to the seminal multiple resource model. We have also made progress in development of stress and workload theory by refining the fundamental questions to be resolved regarding the cognitive resources construct that underlies current stress theory. These efforts, which resulted in four book chapters, one of which is currently in press, will help guide future research on stress and workload associated with human performance. We have made substantive theoretical progress on the question of adaptive interfaces for stressful operations. In particular, we have presented a number of critical insights about the efficacy of design and the possibility of optimal design. These developments have helped characterize and strengthen currently tested empirical propositions that promise to save significant dollars against a vision of endless, iterative search for design improvement. These achievements were facilitated by naturalistic field observations and evaluations. The opportunity to conduct real-world test and evaluation on stressed soldiers and police officers accords with our theoretical proposition that a fully articulated theory can only emerge from an understanding of actual operations. Our experiments on combined physical and cognitive effort have resulted in a surprising and important discovery for this MURI. We call this the 'table-top' tolerance limit phenomenon. This phenomenon indicates that even when operational stress levels are extremely high, the performance remains relatively constant to a point of sudden incipient failure. This finding has significant implications for military operations. Identifying symptomatic characteristics at this "edge of failure" (e.g., progressive increase in performance variability) can provide a predictive marker that enables the identification of impaired soldiers. This may well prove to be a crucial theoretical and practical discovery. Our empirical investigations of workload and shooting performance have enhanced the understanding of the workload-performance relations in real-world tasks, and how personality differences affect these relations. A manuscript based on this work has been submitted for publication

and is contained in this final report. These observations have crucial importance for both military personnel selection and military personnel training. These results support our contention that consideration of individual differences is an important component of research on stress and performance. We also were able to establish theoretical frameworks and experimental paradigms for future efforts to understand the ways in which team performance under stress is not simply the product of individual contributions. Specifically, members of our MURI group were able to show how workload sharing and task division and reallocation by team units acts to diffuse unacceptably high levels of demand, and the importance of task interdependence on team communication and performance. This insight has been used as a basis for designs in adaptive interfaces for stress in which demand may be shared between exposed soldiers and support personnel in remote locations. In addition, the application of neuroergonomic measures to understanding the stress-performance relationship present the next generation level of conceptions for human-machine interaction. We have also found that stress influences mutual trust levels in combined decision-making conditions, and that tactile stimulation can mitigate this problem. Finally, empirical research on Fuzzy Signal Detection theory has supported the contention that this model can be employed to better understand and quantify detection performance when the stimulus dimensions inherently contain high levels of uncertainty. However, constraining operator response to yes/no decisions can reduce sensitivity when the stimuli to be inspected are a fuzzy set.

Capstone Project: Performance Under Stress (P.A. Hancock & J.L. Szalma, Eds).

We are in the process of preparing a text to be published by Ashgate that will describe the research programs funded under this MURI (see Hancock, & Szalma, in press). This text will serve as both summary of research and contribution to improvement in stress theory, and as a summary of tools and methodologies that can be used to assess soldier performance under stress from the team level down to the physiological response of the individual soldier.

MURI Transitions

There have been a series of direct impact or facilitation of contacts and developments with the US Army, including subsequent funding which has been obtained as a direct result of MURI original support. These are specified below.

- Fort Lewis, (Washington State). Associated Army Program: Land Warrior (LW) – Investigation of Scalability of Display issues within LW: Contact

Individuals LTC Cummings, PM Land Warrior, COL Hansen, PM SWAR. Work Funded directly by ARL-HRED.

- Fort Benning, (Georgia). Associated Army program: Future Force Warrior (FFW) and the implementation of Multi-modal displays; tactile displays: Contact Individual: Dr. Elizabeth Redden, ARL/HRED, CSM® Sam Spears, FFW. Work Funded Directly by ARL-HRED.

- Detroit, (Michigan) Associated Army Program: TARDEC Robotic control, visual displays and tactile interfaces. Contact Individual: Terrance Tierney, TARDEC, Mr. Michael Barnes, ARL-HRED. Work Funded Directly by ARL-HRED.

- Fort Benning, (Georgia). Associated Army program: Tactile communication Field Studies. Contact Individual: Dr. Elizabeth Redden, ARL/HRED, Work Funded Directly by ARL-HRED.

- West Point, (New York) Associated Army Initiative with Lockheed Martin USMA Combat Leader Environment. (In association with GEN® Robert Scales, Colgen). Contact Individual: Dr. Michael Matthews, Program Manager Engineering Psychology, USMA. Funding Application in Progress.

- Fort Knox, (Kentucky) Associated Army Program Usability Analysis of Robotic Controls and Displays. Contact Individual: Terrance Tierney, TARDEC, Dr. Keryl Cosenzo, ARL-HRED. Funded Directly by ARL-HRED.

- Fort Chaffee, (Arkansas). Associated Army Program, Special Forces SERE (Survival, Evasion, Resistance and Escape) Group. Associated Researcher Dr. Andy Morgan. Contact Individual: Mr. Brian Pardoe.

- Camp Pendleton (California) Associated Marine Group. Instructors Live Fire Group. Contact Individual: Dr. S Murray.

- Fort Bragg, (Arkansas). Associated Army Program, Field Exercise for Active Soldier Groups. Field Testing. Contact : SERE Group

- A Multimodal Approach to Unmanned Vehicle Operations in Dynamic Environments. Principal Investigator R. Gilson, Co-Principal Investigator, P.A. Hancock, SUS Additional Funding, From ARL through IST

- Adaptive Automation in Human-Agent Teams: The Effects of Role Change, Task Demand, & Operator Characteristics. Principal Investigator J. Szalma, Co-Principal Investigator, P.A. Hancock, SUS Additional From ARL through IST

- Component Improvement for Soldier-Centered Tools: Meta-Analysis of Thermal Effects on Soldier Performance. This information was used as an input to the IMPRINT Soldier Performance Model. Funding from ARL-HRED, Contact Monitor Mr. John Lockett.

- Component Improvement for Soldier-Centered Tools: Meta-Analysis of Noise Effects on Soldier Performance. This information was used as an input to the IMPRINT Soldier Performance Model. Funding from ARL-HRED, Contact Monitor Mr. John Lockett.

- Component Improvement for Soldier-Centered Tools: Meta-Analysis of Vibration Effects on Soldier Performance. This information was used as an input to the IMPRINT Soldier Performance Model. Funding from ARL-HRED, Contact Monitor Mr. John Lockett.

- Component Improvement for Soldier-Centered Tools: Meta-Analysis of Time Pressure Effects on Soldier Performance. This information was used as an input to the IMPRINT Soldier Performance Model. Funding from ARL-HRED, Contact Monitor Mr. John Lockett.

- Scalable Interfaces: Evaluation of Specific Interface Design for Soldier-UAV Interaction. Funding from ARL-HRED, Contact Monitor Mr. Mike Barnes.

- Scalable Interfaces for Enhanced Soldier Performance: Evaluation of Interface Design Parameters for Soldier-Technology Interaction. Associated Field testing, Ft. Benning, Georgia. Funding from ARL-HRED, Contact Monitor Mr. Mike Barnes. Ft. Benning Contact, Dr. E. Redden.

- Human-Robot Interaction Component of the "ECATT/MR STO Research and Testbed Product Integration" From the Research and Development and

Evaluation Command, (REDCOM) United States Army, Under Contract No. N61339-05-C-0046. Contact Monitor Mr. Mike Barnes.

Students Supported

At UCF MIT^2 Laboratory alone, there were a total of eight Post-Doctoral Researchers supported. Other contributing Laboratories added at least a further 3 individuals to this total.

In terms of Graduate Students MIT^2 Laboratory supported more than Graduate Students of which five have now completed the Ph.D and have gone on to work in Human Factors related position. That number is expected to grow in the coming year by at least another four students. In allied Laboratories at least four further students have completed the Ph.D. The number of Master's completing students is in proportion to this upper level effort.

In terms of Undergraduate students, the MIT^2 Laboratory has had at least seven full-time individuals, many of whom have now proceeded to Graduate School. Concomitant numbers have been achieved in other affiliated Laboratories. Thus, in terms of the training and pedagogical facet of the MURI endeavor, the MURI-OPUS program has been especially successful and we will see the fruit of this investment for many years to come in on-going research programs in this vital area.

References

Hancock, P.A., & Szalma, J.L. (Eds.). (2007). *Performance under stress*. Ashgate Publishing, London: England, in press.

Hancock, P.A., & Warm, J.S., (1989). A dynamic model of stress and sustained attention. *Human Factors*, **31**, 519-537.

Scales, R.H. (2006). *Clausewitz and World War IV*. Armed Forces Journal, July.